Fundamentals of Media Security

WeiQi Yan; Jonathan Weir

WeiQi Yan & Jonathan Weir

Fundamentals of Media Security

Fundamentals of Media Security

Contents

Bibliography

List of Figures

List of Tables

Chapter 1

Introduction

With the development of computing technologies, digital media security becomes an ever increasing issue of daily life. As computer software and hardware evolves and as more of it becomes accessible as an Internet application or service, information security has gone from initially protecting digital images, into the area of multimedia. This area of multimedia includes computer graphics, images, digital video and digital audio. As the Internet expands, so does the significance of multimedia security. This chapter will cover some of the basic knowledge required and what should be considered when discussing media security. Analysis of the theoretical background of these issues is given, which is used to guide the practical applications.

1.1   Overview

Computer security with regard to system security and network security has been broadly studied. With the development of multimedia processing and applications, multiple types of media security have been broadly taken into consideration. The type of media includes audio, video, image, text, web pages and graphics.

From a hardware perspective, in order to protect the copyright of Intelligent Property (IP), special fonts were written into the ROM on microchips so that fraudulent chips could be identified. If these marks are removed, the fragile microchip will no long function. This technique has been employed even today, the microchip protections are still regarded as one of the best ways to protect information pertaining to that chip. Another typical software protection is a "serial number", which attempts to block illicit copying, along with online Internet registration of that software.

Visually, logos have been adopted as a type of visible watermark, used frequently within the television industry. These marks are used to denote content ownership.

Nowadays, secure protection approaches have been embedded into the commercial products such as computer games, digital music, digital video, engineers drawings and many other forms of digital media and documents. Due to the rapid expansion of the amount of digital content available, illegal usage of these products can have many detrimental effects. It cannot be denied that media security will play a vital role in the impact multimedia products will have over a long period of time.

Although the research in the area of media security has made great progress in the past ten years, there are still many problems with existing products and additionally, during that time, many more problems, previously not considered have presented themselves. The robustness and capabilities of the existing security models have not been sufficiently investigated or updated to handle these new problems.

In this book, the research in the area of media security will be presented from a mathematical point of view, starting from the basic principles of cryptography, and how they can be used for a much wider application in terms of media security. The research areas that this book covers are presented within the next section.

1.2   Research Areas of Media Security

Media security is one of the family members of information security. Media security including information hiding is built upon cryptography and multimedia processing. The primary research topics covered within this book include steganography, watermarking, scrambling, authentication and forensics, and surveillance. This type of research is commonly employed in digital commercial products for tasks such as evidence verification and authentication [4]. The following paragraphs give brief details of these techniques, making it clear what will be covered in the following chapters. These paragraphs correspond to the outline of this books content.

Steganography, namely information hiding refers to hiding secrets in other media. The host media should remain fully complete and contain all original data. The hidden media should be embedded with a certain degree of robustness so that the secret can resist certain attacks, but must also be fully recoverable after such attacks or manipulations in order to ensure the secret can be correctly extracted.

Watermarking is a typical problem in information hiding. After a secret is embedded into some type of media, that media can then be inspected or tampered with. Even after this inspection and tampering, the hidden information, in the form of a watermark, should still be intact such that it can be extracted as evidence to prove the ownership. A digital watermark usually refers to a tiny binary sequence. An example of a typical well known watermarking algorithm would be the LSB (Least Significant Bits) algorithm. The key issues in information hiding and watermarking

are watermark capacity and robustness. Fragile watermarking is another technique that can be applied in information protection. Logos on videos and images are a visual watermark.

Scrambling is primarily employed to break the coherence within the temporal and spatial domain in order to encrypt a television signal. This is to help protect the content and owners of that content from piracy. The reason why scrambling is researched is that the approaches based on cryptography do not affect the signals which have spatial or temporal coherence. Progressive scrambling is specially designed for multimedia data, such as audio, video, and images. This progressive scrambling can break the coherence at multiple resolutions, like a public key system. Without the keys, the secret cannot be restored even if the algorithms are public. If some of the keys are given, the media, in some degraded quality can be partially decoded and viewed. If all the keys are provided, the media can be fully decoded and displayed in perfect, original quality.

Media authentication and forensics are an emerging problem in information verification. With the rapid development of Internet applications and digital devices, such as digital cameras, video cameras, microphones and scanners, digital forensics have been playing a pivotal role in information processing. Some people hope real photos are fake while some expect the fake photos are real. The responsibility of media authentication and forensics is to determine the truth about particular events which have been photographed or videoed. Various techniques are used in analyzing media in order to determine whether or not it has been tampered with.
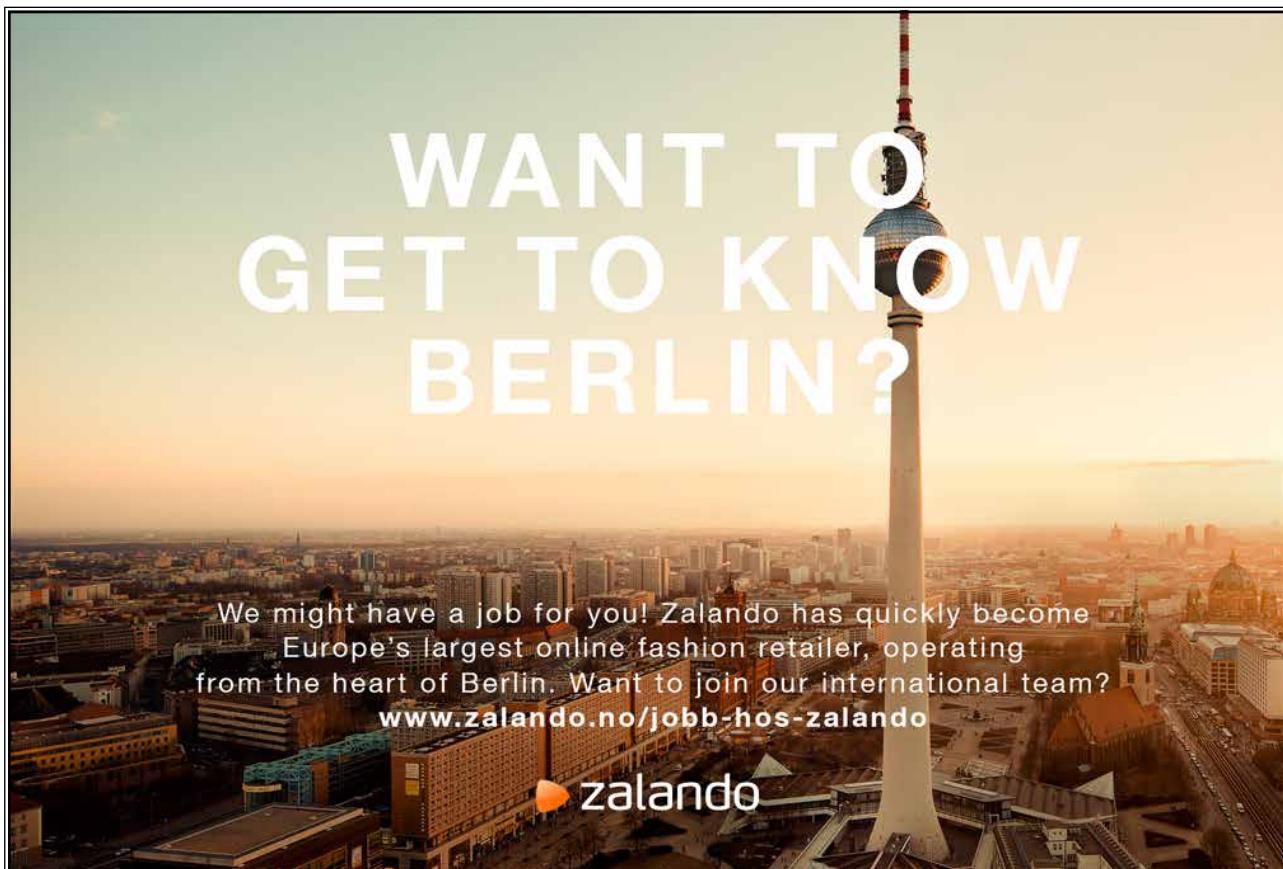
Video surveillance is a hybrid of computer vision and media security. The main problems in video surveillance are: capturing the video and efficiently storing it, online camera control, visual search and event mining using the stored data. Due to the expansive nature of recorded video, processing the shear volume of material effectively is another primary concern. Therefore, the main techniques required in this area are object detection and tracking, trajectory analysis, occlusion analysis, visual search and multiple camera coordination. Depending on the application, these techniques can be used to pinpoint certain activities that the surveillance system may be responsible for.

There exists many unsolved problems within the media security domain. Although all of these problems cannot be listed due to space limitations, it is hoped that through the use of this book, they may present themselves to the reader. As the practical requirements for this type of media security work increase, these techniques will need to be improved and enhanced to handle more specific cases. Before an examination of each of the techniques are presented in detail, a brief overview of the evaluation criteria used for analyzing said techniques is given. This criteria should be considered when working in area of media security.

1.3   Criteria for Media Evaluation

In media security, if the tampered data and the original data have significant differences, then an objective evaluation method is not required. It will be apparent to an observer that the image has been altered by some degree. The quality of multimedia data can be measured subjectively by a number of methods, such as, understandable or perceivable difference, or degree of resemblance that occurs.

The understandable or perceivable difference corresponds to the information capacity that a human or machine can obtain by observing the original and altered media. The degree of resemblance refers to the difference between the original standard data and the new target data. The criteria for media evaluation includes two main aspects: subjective and objective criteria [5]. The following sections present mathematical inspections of each type of criteria in order cement a solid base upon which to start.

1.3.1   Subjective Criteria

Subjective evaluation is performed by the human perceptual system, it is closely related to media quality but also the characteristics and conditions of the observers. Subjective evaluation includes absolute evaluation and relative evaluation. It is calculated by:

$$C = \frac{\sum_{i=1}^{k} n_i c_i}{\sum_{i=1}^{k} n_i} \qquad (1.1)$$

where c_i is the score of class i, n_i is the number of people evaluating the image that belongs to class i [5]. That is, given a group of people, how perceivable are the changes that have been made to the original media. Results may vary in this type of evaluation due to the nature of sight in that some people's vision operates in a much less efficient way to anothers. Age may also have an adverse effect on this type of study.

1.3.2   Objective Criteria

Due to the limitations that occur when subjective criteria is used for examining media, objective criteria was introduced. An objective and more scientific way of examining and testing for differences between a set of data is to use objective criteria methods. These objective methods help to quantify senses and perceptions that may be somewhat more difficult to describe in terms of computing on a finite scale. These objective methods attempt to bridge the gap between what a group of users perceive and what is actually happening.

Suppose the intensity of a pixel within an image I at coordinate (i, j) is $f(i, j)$, $0 \leq f(i, j) \leq 255$; and suppose the pixel intensity within another image I' at coordinate (i, j) is $f'(i, j)$, $0 \leq f'(i, j) \leq 255$, the difference between image I and image I' may be described using the following objective criteria and methods:

1. Definition of Reality [5]:

$$\varepsilon = \frac{\sum_{j=1}^{M} \sum_{k=1}^{N} [\sigma\{f(i,j) - \sigma\{\hat{f}(j,k)\}\}]^2}{\sum_{j=1}^{M} \sum_{k=1}^{N} [\sigma\{f(i,j)\}]^2} \qquad (1.2)$$

   where $\sigma\{\cdot\}$ is the probability expectation [6].

2. Peak Signal to Noise Ratio (PSNR) is defined as [7]:

$$PSNR = 10 \log \left(\frac{m \cdot n \cdot P^2}{RMS} \right) \qquad (1.3)$$

   where P is the peak of the signals, m and n are the horizontal and vertical resolution, RMS (Root Mean Square) is:

$$RMS = \int_{l^2} (f(x,y) - f'(x,y))^2 dx dy \qquad (1.4)$$

For the discrete case:

$$RMS = \sum_{i=1}^{n} \sum_{j=1}^{m} (f(i,j) - f'(i,j))^2 \qquad (1.5)$$

3. Energy signal noise ratio is defined as [7]:

$$SNR = \frac{S^2(x,y)}{\varepsilon\{N^2(x,y)\}} \qquad (1.6)$$

where $S(x,y)$ is the energy signal, $N(x,y)$ is the energy of noise, $\varepsilon\{\cdot\}$ is the expectation.

The PSNR metric is one of the most common objective methods used, specifically within the areas of watermarking and steganography. After watermark or information embedding, high PSNR values of the altered/watermarked image correspond to high similarity against the original. Further examples are provided later in this chapter which show the capacity of certain types of images for information hiding.

1.3.3   Distance Measure on Various Color Spaces

The distance measure between two colours is another objective method used for analyzing a set of images. It allows one to quantify the difference between two pixel values that would otherwise be subjective if described by a person.

Firstly, the distance measure on YIQ colour space is examined. The transform between RGB and YIQ is given by Eq. (1.7), where YIQ is the brightness, saturation and chrome.

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & 0.523 & 0.312 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1.7}$$

Thus, given two images, the difference can be measured by:

$$RMS_{YIQ} = \frac{\sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (Y_{ij} - Y'_{ij})^2 + (I_{ij} - I'_{ij})^2 + (Q_{ij} - Q'_{ij})^2}}{m \cdot n} \tag{1.8}$$

where (Y_{ij}, I_{ij}, Q_{ij}) and $(Y'_{ij}, I'_{ij}, Q'_{ij})$ belong to two different resolution images, m, n are the resolution in the vertical and horizontal direction.

The perception of two images $L \cdot a \cdot b$ from CIE is equal when two images are computed. The steps required to convert from an RGB colour space to the $L \cdot a \cdot b$ colour space are described as follows:

Firstly, complete the conversion from RGB space to CIE XYZ space:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.431 & 0.342 & 0.178 \\ 0.222 & 0.707 & 0.071 \\ 0.020 & 0.130 & 0.939 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1.9}$$

Secondly, from XYZ space to $L \cdot a \cdot b$ space:

$$L = \begin{cases} 116 \left(\frac{Y}{Y'} \right)^{\frac{1}{3}} - 16 & \text{if } \frac{Y}{Y'} > 0.008856 \\ 903.3 \frac{Y}{Y'} & \text{if } \frac{Y}{Y'} \leq 0.008856 \end{cases} \tag{1.10}$$

$$a = 500 \left[f \left[\frac{X}{X'} \right] - f \left[\frac{X}{X'} \right] \right] \tag{1.11}$$

$$b = 500 \left[f \left[\frac{Y}{Y'} \right] - f \left[\frac{Z}{Z'} \right] \right] \tag{1.12}$$

where

$$f\left(\frac{V}{V_n}\right) = \begin{cases} \left(\frac{V}{V'}\right)^{\frac{1}{3}} & \frac{V}{V'} > 0.008856 \\ 7.787\frac{Y}{Y'} + \frac{16}{116} & \frac{V}{V'} \leq 0.008856 \end{cases} \tag{1.13}$$

and (X', Y', Z') is the reference, L stands for the luminance, a is the degree from green to red, b is the degree from blue to yellow. Therefore, in CIE $L \cdot a \cdot b$, the distance between two images is defined as:

$$RMS_{Lab} = \frac{\sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (L_{ij} + L'_{ij})^2 + (a_{ij} - a'_{ij})^2 + (b_{ij} - b'_{ij})^2}}{m \cdot n} \tag{1.14}$$

where (L_{ij}, a_{ij}, b_{ij}) and $(L'_{ij}, a'_{ij}, b'_{ij})$ belong to different images having the same resolution, m, n is the resolution on the vertical and horizontal direction [7].

1.3.4   Comparison Among Variant Measurement Approaches

The above approaches are divided into subjective and objective evaluation criteria. There are two types of object evaluation criteria: The 1st evaluation is based on comparing two images in the spatial domain using the L_2-norm in an Euclidean space by computing the average values. The comprising result reflects the physical differences between two images, however it is a weak comparison based purely on content of an image. The 2nd is a comparison based on statistics. This approach attempts to solve the resemblance problem from a statistical point of view. It supports invariant attributes such as rotation and translation of an image. Therefore, the later approaches are reasonable when measuring the quality of images. The following sections consider this statistical approach from an mathematical viewpoint.

1.4   Mathematical Models for Information Hiding

When information hiding in terms of images is discussed, the first question is: Based on an image's bandwidth and signal to noise ratio (SNR), how much information can be hidden within the image? That question can be answered by following several facets:

1.4.1   Mathematical Models Based on Colour Space

Given $f(i, j)$ of image **A** and $g(i, j)$ of image **B**, (i, j) is the coordinate on an image. According to these discrete properties, $f(i, j)$ and $g(i, j)$ may be represented by:

$$f(x, y) = \sum_{i=0}^{7} b_i \cdot 2^i \text{ and } g(x, y) = \sum_{i=0}^{7} b'_i \cdot 2^i \qquad (1.15)$$

where $b_i = 0, 1$, $b'_i = 0, 1$, $i = 0, 1, \ldots, 7$.

$$f(x, y) = \sum_{i=4}^{7} b_i \cdot 2^i + O(2^4) \qquad (1.16)$$

$$g(x, y) = \sum_{i=4}^{7} b'_i \cdot 2^i + O(2^4) \qquad (1.17)$$

Using this, the following function can be constructed:

$$h(x, y) = \sum_{i=4}^{7} b_i \cdot 2^i + \sum_{i=4}^{7} b'_i \cdot 2^{i-4} \qquad (1.18)$$

Using bit-operations:

$$f'(x, y) = \sum_{i=4}^{7} b_i \cdot 2^i \text{ and } g'(x, y) = \sum_{i=0}^{3} b'_i \cdot 2^{i+4} \qquad (1.19)$$

The error of the image restoration can be calculated using:

$$|f(x, y) - f'(x, y)| \le 16 \text{ and } |g(x, y) - g'(x, y)| \le 16 \qquad (1.20)$$

Since the humans visual systems (HVS) sensitivity can be measured on a logarithmic scale, very bright pixels can have a difference of 16 continuous tones, the HVS cannot detect these differences. Therefore, the pixels that have 16 continuous degrees of difference will not be perceptible with 256 colour tones. Thus, the information within this range will be hidden. This simple principle is widely employed in information hiding. That is, if it cannot be seen by humans, it is more than likely suitable as a base for information hiding.

1.4.2   Mathematical Models Based on PSNR

Given a public image which contains a secret, suppose an 8-bit grayscale image is used, the definition of PSNR (Peak Signal Noise Ratio) is:

$$PSNR = 10 \log \left[\frac{M \times N \times 255^2}{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (f(i,j) - g(i,j))^2} \right] \qquad (1.21)$$

where $f(i,j)$ is the original greyscale image, $g(i,j)$ is the pixel value from the reference image (the image that contains the secret), (i,j) represents the coordinate of the pixel on each image.

From experiments, it is known that if two images have minimal differences, their PSNR should satisfy:

$$PSNR \geq 28 \qquad (1.22)$$

Table 1.1 and Table 1.2 are the comparison after using the simplest LSB watermarking algorithms for different image types. Table 1.1 corresponds to a greyscale image while Table 1.2 represents the results from an RGB image.

Table 1.1: A table comparing the PSNR after different bit changes on a greyscale image.

The bit to be changed	Corresponding PSNR
1st	38.62
2nd	36.12
3rd	30.10
4th	24.08
5th	18.06
6th	12.04
7th	6.02

From Table 1.1 the following conclusion can be drawn: for an 8-bit grayscale image, the four lowest bits can be employed to hide information. Correspondingly, from Table 1.2, for a 24-bit color image, the three lowest bits can be employed for information hiding.

1.4.3   Channel Model of Information Hiding

Information theory is generally considered as being founded in 1948 by Claude Shannon in his seminal work, "A Mathematical Theory of Communication" [8]. The

Table 1.2: A table comparing the PSNR after different bit changes on an RGB image.

The bit to be changed	Corresponding PSNR
1st	47.71
2nd	37.37
3rd	27.09
4th	21.35
5th	19.31
6th	13.29
7th	7.26

central paradigm of classical information theory is the engineering problem of the transmission of information over a noisy channel. The channel capacity can be approached by using appropriate encoding and decoding systems.

In information theory, there are two important concepts: entropy and degree of security, they are the important criteria for secure communication and information measurement. Any secure communication system can be evaluated using these two basic concepts, and they are quite reliable for new system design.

Generally, if there are n messages, their probabilities are $P_1(x), P_2(x), \ldots, P_n(x)$ respectively, then it can be proven that the average information capacity of each message is:

$$H = -\sum_{i=1}^{n} P_i(x) \log_2 P_i(x) \tag{1.23}$$

Eq. 1.23 reflects the entropy.

Smith and Comiskey [9] presented a communication model of information hiding from information theory based on digital images [10]. They synthetically take capacity, robustness and imperceptible of information hiding into consideration. The cover image is thought as 2-dimensional commutation channel with limited bandwidth having average noise energy. The secret message is transmitted from this channel. Starting from this simple model, based on Nyquist and Shannon theorem, the capacity of this communication channel in the given bandwidth can be calculated, and the inferior boundary of S/N to obtain the capacity C:

$$C = W \log_2^{\left(1+\frac{S}{wn_0}\right)} \tag{1.24}$$

When bandwidth w tends towards infinity:

$$\lim_{w\to\infty} C = \lim_{w\to\infty} w \log_2^{\left(1+\frac{S}{wn_0}\right)} \approx 1.44 \frac{S}{n_0} \tag{1.25}$$

This indicates that when the bandwidth tends to infinity, the channel capacity will become a constant instead of infinity.

Ramukumar et al. [11] improved the above model by designing a model with increased noise, and analyzed the performance of the system. In the system, there are two noise sources: noise I is introduced by modulation with the original images while P is introduced by signal processing including compression and decompression. S is the secret transmitting through the noisy channel. In information detection, if the original image is used as a reference in information hiding, the channel contains noise P only. $X \to [f_X(x), \delta_x^2]$ is the information to be transmitted; $Z \to [f_Z(z), \delta_z^2]$ is the noise in the channel; $Y \to [f_Y(y), \delta_y^2]$ is the information received from the output port. Suppose X and Z are independent on each other, the capacity of this noisy channel C in information theory is:

$$C = \max_{f_X(x)} I(X,Y) = \max_{f_X(x)}(h(y) - h(Y|X)) = \max_{f_X(x)}(h(y) - h(z)) \tag{1.26}$$

Where $I(X, Y)$ is the information of X and Y. Therefore, given a system with sufficient noise and input, the entropy of Y can be obtained. Hence, the capacity of information is:

$$C_h = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_S^2}{\sigma_P^2 + \sigma_{ig}^2} \right) \tag{1.27}$$

where σ_S^2 is the energy of hidden information, σ_{ig}^2 is the energy of modulation noise, σ_P^2 is the energy of signal processing.

By simply analyzing the perceptual and impartial information, the capacity of communication channels in information hiding can be understood. If the imperceptible information in the media is the real communication provider, and Smith's basic scheme is adopted, from Eq. 1.27, the following is obtained:

$$C_h = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_S^2}{\gamma \cdot \sigma_{ig}^2 + \sigma_P^2} \right), 0 \leq \gamma \leq 1 \tag{1.28}$$

In Eq. 1.28, the factor of image energy γ is introduced to show that the noise is from part of the image not the whole image. If modulation and demodulation are considered as a pair of reversible transformations, the reversible degree reflects the perceptual degree of imperceptible information. If $\gamma = 0$, it means the sender owns the original information and the receiver gets a clear piece of information.

1.5 Theoretical Security Model for Information Hiding

In security communication, the space spanned by cipher text set X is called the cipher space. The space spanned by encrypted set Y is called encrypted space while the key set $K = \{k_i, i = 1, 2, \ldots, r\}$ is called key space. The encryption transformation T_k is the mapping from the cypher space to the encrypted space, the decryption transformation T_k^{-1} is the mapping from the encrypted space to the cypher space, k represents the elements in the key space K. If $P(x, k)$ is the joint probability of X and K, then the definition of a secure system is: $C \triangleq \{X, K, Y; P(x, k_i, y)\}$, or C is a secure system defined by set X, Y, K and distribution $P(x, k_i, y)$. T_k maps $x \in X$ to $y \in Y$, therefore,

$$P(x, k_i, y) = P(x_i, k) T(y|x, k_i) \tag{1.29}$$

where

$$T(y|x, k_i) = \begin{cases} 1, & y = T_k(x) \\ 0, & \text{otherwise} \end{cases} \tag{1.30}$$

The principle when measuring whether a communication channel is secure or not includes the adopted degree of security, size of keys, complexity of encryption and decryption along with error spreading.

The secure model of information hiding is built on a test. Suppose Alice sends a message to Bob, attackers should judge whether it includes hidden information. If it does, how much is hidden and what is it? The famous hypothesis in information theory and mathematical statistics should be used. Now, a Bayesian approach is used along with the average entropy $D(P_C|P_S)$ to quantize the security between public cipher text P_C and pseudo cipher text P_S. If the entropy distribution of public cipher text and pseudo cipher text is $D(P_C|P_S) = 0$, then the attacker cannot get the hidden context, unless the secret was already known beforehand.

In this modal, the transmission procedure is regarded as a random procedure. During a given period of time, the encrypted results are different from other encrypted results, therefore, it is secure. In terms of validation and security, the key problem in information hiding is robustness and accuracy of a system.

The potentiality of set S is noted as $|S|$. The entropy of x and probability P_X is defined as:

$$H(x) = -\sum_{x \in X} P_X(x) \log P_X(x) \tag{1.31}$$

The condition entropy of X given Y is $H(X|Y) = -\sum_{y \in Y} P_Y(y) H(X|Y = y)$, where $H(X|Y = y)$ is the entropy of conditional distribution $P_{X|Y=y}$.

Definition 1.5.1 (Security). *If $D(P_C||P_S) \leq \varepsilon$, then the system consisting of cover information C and secret S are E-passive secure. If $\varepsilon = 0$, the system is completely secure [10].*

Theorem 1.5.2 (Detected Probability). *In an E-passive secure system, the probability that an attacker cannot detect the secret α and the possibility that the wrong secret β is obtained, satisfies:*

$$d(\alpha, \beta) \leq \varepsilon \tag{1.32}$$

$d(\cdot)$ is the relative entropy, if $\alpha = 0$, then

$$\beta > 2^{-\varepsilon} \tag{1.33}$$

In a secure system, $D(P_C||P_S) = 0$, therefore, $P_C = P_S$ [10, 11].

Theorem 1.5.3 (Passive Security). *Security of the system mentioned above which only uses one bit in information hiding is passive secure to the attacker [10, 11].*

Theorem 1.5.4 (Security of Compression System). *Information compression is a stable procedure, if the compression is an E-passive secure system, then the compressed data is at least E-passive secure [10, 11].*

1.6 Summary

This chapter primarily explains the definition of media security and emphasizes the type of media security that will be covered within the remainder of this book. Also discussed were several mathematical models used within the area of information hiding and the corresponding evaluation criteria used for analyzing such digital media. Finally, the capacity issues pertinent to information hiding were discussed, and the relevant mathematical models presented to help illustrate what must be considered when designing such systems. This chapter will have provided the reader with a good base from which to start. The following chapters take each of the specific topics into considering and provide a wider scope for each of the methods discussed.

Chapter 2

Steganography

Steganography, namely information hiding refers to hiding secrets in other media. The host media should contain the majority of its original content. The hidden media should be embedded with certain robustness so that the secret can resist certain attacks. This ensures the secret can be correctly extracted. This chapter, introduces the history, status and trends in information hiding in terms of steganography. Using digital multimedia data such as video and audio as the host media, a comparison and analysis for some well known information hiding algorithms is presented.

2.1   Overview

The word steganography is of Greek origin and means "covered", or "hidden writing". Its ancient origins can be traced back to 440 B.C. Herodotus mentions two examples of steganography in The Histories of Herodotus. Demaratus sent a warning about a forthcoming attack to Greece by writing it on a wooden panel and covering it in wax. Wax tablets were in common use then as re-usable writing surfaces, sometimes used for shorthand. Another ancient example is that of Histiaeus, who shaved the head of his most trusted slave and tattooed a message on it. After his hair had grown the message was hidden. The purpose was to instigate a revolt against the Persians. Later, Johannes Trithemius published the book Steganographia, a treatise on cryptography and steganography disguised as a grimoire. Today, the term steganography includes the concealment of digital information within computer files.

Steganography always adopts hidden messages on paper written in secret inks under other messages or on the blank parts of other messages. During and after World War II, espionage agents used microdots (photographically produced) to send information back and forth. Since the dots were usually extremely small (the size of a period produced by a typewriter or even smaller), the stegotext was whatever the dot

was hidden within. If the dots were concealed within a letter or an address, they were usually inside some alphabetic characters. If under a postage stamp, it was the presence of the stamp. The problem with the WWII microdots was that they needed to be embedded in the paper, and covered with an adhesive (such as collodion), which could be detected by holding a suspected paper up to a light and viewing it almost edge on. The embedded microdot would reflect light differently than the paper [12].

Steganography also adopts "zero encryption". Zero encryption corresponds to hiding secret messages inside other forms of information, the information can be filtered, but the hidden secret cannot be determined from the filtering process. The following example illustrates this:

> "News Eight Weather: Tonight Increasing Snow. Unexpected Precipitation Smothers Eastern Towns. Be Extremely Cautious And Use Snowtires Especially Heading East. The Highways Are Not Knowingly Slippery. Highway Evacuation Is Suspected. Police Report Emergency Situations In Downtown Ending Near Tuesday."

If the first letter of each word is extracted, the following message can be obtained:

> "Newt is upset because he thinks he is President."

The following message was sent out from German spies in World War II:

> "Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on byproducts, ejecting suets and vegetable oils."

If the second letter from each sentence is extracted, the following message emerges:

> "Pershing sails from NY June 1."

Steganography refers to hiding secrets in public messages. Compared with cryptography, it allows potential adversaries to detect and intercept, even modify the public secret, while, at the same time does not violate any secure limitation.

The aim of information hiding is to hide information inside innocent information in order to alleviate any suspicion that some type of information may be hidden within the innocent data. Any adversary monitoring the innocent information should have a hard time proving the existence of any hidden information. The cover media used to hide the secret message may be an image, audio, video, cipher text or any

other form of digital media. The secret message can be text, another file or binary streams. The secret and the cover media merge together to form an information carrier. Information hiding may need a pseudo key or password as the attached information. When a secret has been hidden in protected media, the media is normally referred to as pseudo media. Generally,

Protected info + Embedded Info + Pseudo Key = Pseudo Info

The attempt to detect and break a steganographically encoded packages is called steganalysis. The simplest method to detect modified files, however, is to compare them to the originals. To detect information being moved through the graphics on a website, for example, an analyst can maintain known-clean copies of these materials and compare them against the current content on the site. The differences (assuming the carrier is the same) will compose of a payload. In information hiding, the cover information, pseudo information and key are employed to compare the new information. Steganalysis is the counterpart of Cryptanalysis, the knowledge of Cryptanalysis can be employed in steganalysis [12].

Information hiding provides imperceptible information encryption, this helps to reduce the likelihood of the hidden information from being detected. If the secret can be hidden within public images, the signal transmission will be very much safe, and will have a lower likelihood of being attacked or monitored [13].

The tag of a confidential document for indexing purpose plays a pivotal role when comparing it against the physical document. If the confidential document is physically changed or has marks placed on it, that will be very obvious and dangerous. However, without this tag, the documents become very hard to manage. The solution requires hiding the tags within the document so that the tags stay together with the data. This requires an algorithm to extract these tags precisely and swiftly.

Information hiding is closely related to the Human Perceptual System. This means that if a piece of data is to be used as a cover for some hidden data, the important parts of the original data should not be censored or obscured significantly so that the human perceptual system has difficulty in recognizing what the original should be.

On the other hand, information hiding should have robust attributes, therefore, even if the interceptor already knows that a piece of media contains a secret, perceptually, they cannot determine where it may be located and therefore cannot obtain the secret. Based on the requirements of robustness, currently the technology employed in information hiding is presented within the frequency domain. This is because the spatial domain and color domain have apparent patterns, interceptors can break the security using brute force techniques based on statistical analysis. However,

the frequency domain is rather stable, it not only resists the impact due to various noises that may be introduced, but also the complexity of the frequency domain will require more skill and analysis in terms of breaking it.

2.2 Stego Algorithms

Digital media is an ideal secret carrier for information hiding, this is because digital media has a huge volume of redundancy. This attribute provides the potential for hiding information in this cipher space. Therefore, the techniques and skills in digital media processing have been applied to this area. These approaches are based in the spatial and frequency domains. The spatial domain includes bit-operations, a binary stream can be embedded in the Least Significant Bit (LSB) [14]. As the outcomes are based on statistics, these can be used to prove the ownership of a piece of digital media. This approach is suitable for information hiding and is known as the simplest information hiding system.

The mathematical algorithms used for information hiding within the frequency domain include: Discrete Cosine Transformation (DCT), Discrete Wavelet Transformation (DWT) and the Walsh Transformation. These approaches are based on the transformation of an orthogonal function system. These approaches provide an acceptable trade-off between the secret and the robustness of the information carrier for the final secret extraction. Many such approaches are independent of image format and can be converted between lossy and lossless formats [15].
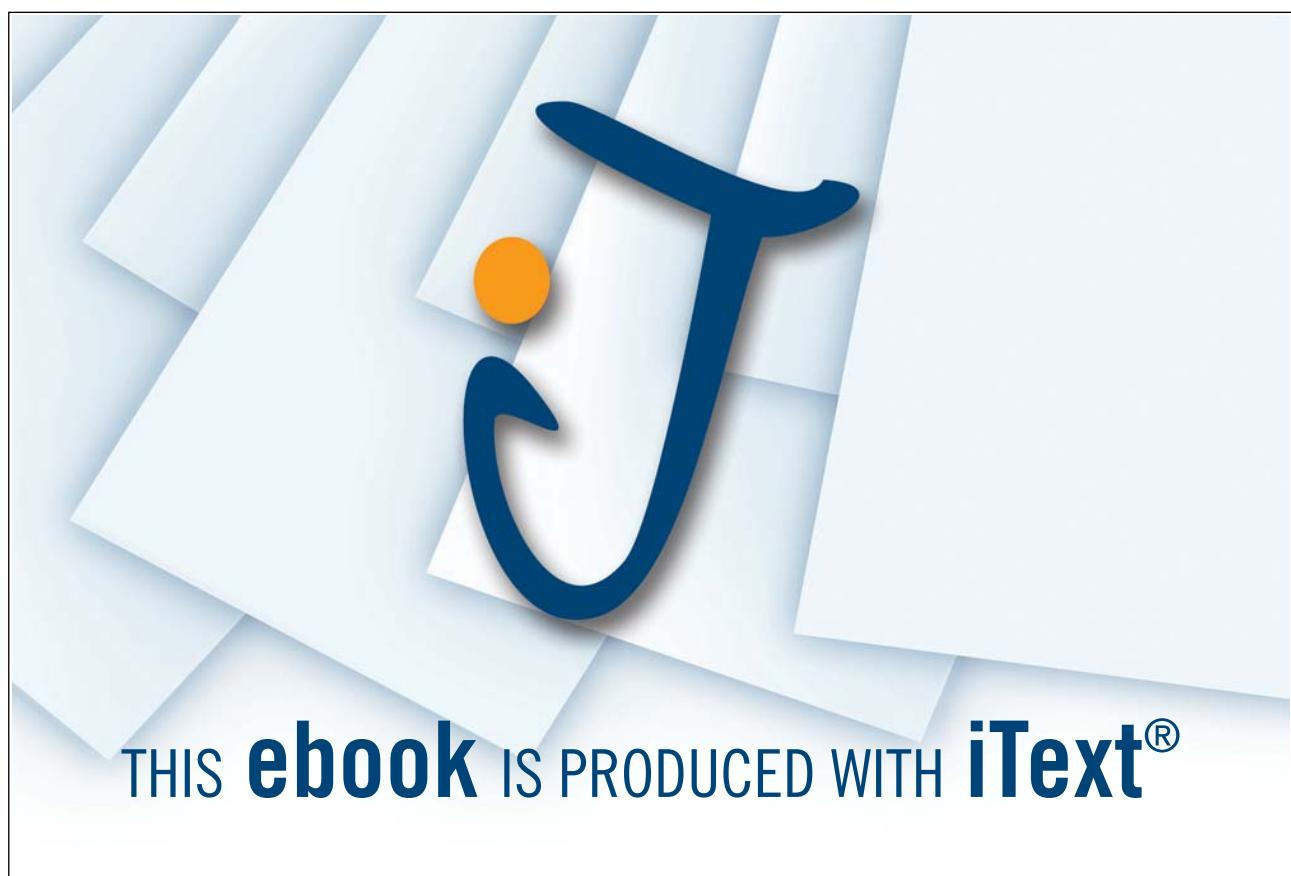
JPEG (Joint Picture Expert Group) and JPEG 2000 utilize DCT and DWT to compress digital images. The compressed data are integers due to quantization and run length coding. However, other approaches have multiple attributes in order to hide information in both the spatial and frequency domains. The algorithms combine these features together. These approaches are helpful to protect the hidden information and use it to resist various attacks, such as rotation, scaling and cropping. A patchwork technique selects multiple regions of an image to hide information, each region includes a tag, even if one tag has been destroyed, other tags still keep the marks so that identification is still possible.

In steganography, the secret is transmitted using public channels, in a public image, to hide the secret message. Digital images and the derived technologies are widely being accepted. In this area, cryptography and forceful encryption are not welcome. People are more interested in sending a secret without arousing any suspicion. This allows the secret to be placed in full public view and can be potentially accessed from anywhere, meanwhile, the secret remains totally imperceptible. Classic commercial applications of information hiding are watermarking in banknotes and digital signatures, these are employed to trace the copyright and ownership of physical and electronic products.

If a watermark exists, and is not perceivable, the customers may assume no watermarks are present, and will have a harder time trying to locate the watermarks if they wish to remove them from commercial products. This is a very challenging topic. This leads into the next section on detecting information hiding.

2.3 Detecting Information Hiding

Information hiding always hides a secret in a huge volume of data, while watermarking hides binary streams in digital media. Thus the watermark digits can be uniformly distributed throughout the media. No matter which scheme is chosen, the effect is imperceptible to the human visual system [16].

In order to evaluate hidden information, it is very important to define some criteria. Generally it is very hard to define a standard image. It is only when the cover image and original image are observed together, that the differences can be pointed out. In an image using color indexing, in order to reduce the reference time, the color is sorted in a specific order, the changes to these colors are minor. Information hiding is very hard to gain from strong contrast images, the minor changes will reflect the images.

The basic way for secret detection in information hiding is to compare the original image and the secret carrying image, and take special note of the distinct details.

A digital signature is a very important piece of information, the signature can be employed to identify the existing hiding information.

Human eyes are very sensitive to the distortion, especially when the secret carrying the media and the original have been taken into comparison. Based on knowledge patterns, similar models may be established, this is very helpful in steganoanalysis. All the information can be automatically detected, the corresponding tools can be employed to detect the secret.

2.4   Bit Operation Based Information Hiding

Generally, the information hiding can be described as following:

Definition 2.4.1 (Information Hiding). *Given digital media A and B, finding the digital media C and D, and reversible transformation F, must satisfy:*

$$|F(\alpha_{i,j}, \beta_{ij}) - \beta_{ij}| \leq \varepsilon \tag{2.1}$$

$$w_{ij} = F^{-1}(\beta_{ij}, F(\alpha_{ij}, \beta_{ij})), |\alpha_{ij} - w_{ij}| \leq \varepsilon \tag{2.2}$$

where $\alpha_{ij} \in A$, $\beta_{ij} \in B$, $w_{ij} \in C$, $F(\alpha_{ij}, \beta_{ij}) \in D$; $0 \leq i \leq m-1$ and $0 \leq j \leq n-1$ are image resolution, $\varepsilon > 0$ is threshold. A is secret image, D is the public image, F is transformation for image information hiding [17].

Information hiding is not a new problem [18], surprisingly, it has origins in old children's games. However, because the rapid development of computer communication, especially due to the swift development of the Internet, the requirements for image information hiding become necessary and urgent. This requirement may be due to the traditional cryptography schemes which have to face the huge volume of data and due to its time consuming nature, is unsuitable for the job of information hiding. New schemes must be presented which allow detection of these methods [19].

Watermarking in actual fact is important in terms of information hiding, however it emphasizes the little tags consisting of random binary numbers. If the hidden information is not minor, this will be an information hiding problem. Usually, the encrypted images will arouse the attention of an attacker. If a secret can be hidden in public media, the media transmission will be very safe, it even can avoid potential attacks. Techniques have been developed which allow watermarking to happen in real-time, specifically when watermarking video [20]. Watermarking will be discussed in further detail in the follow chapter. The research in information hiding requires knowledge from two aspects: mathematical algorithms and validation encoding.

2.5   Mutual Multimedia Information Hiding

This section will introduce the information hiding approaches used for concealing data within other types of non-image related data. The main contents include hiding images in sound, sound in sound along with a number of other approaches.

2.5.1   Images Hidden in Sound[1]

In order to hide images in sound, the wave format is introduced. In order to hide images in sound, the wave file needs to be opened, and combined with the secret data according to the following equations:

$$z = f(i, j) \tag{2.3}$$

$$y = g(i, j) \tag{2.4}$$

Where z is merged data, i is the data of the wave form file, j is the image file, f is the operation which hide images into sound, (u, v) is the coordinates of each pixel, g is the transformation from color image space to one dimensional space.

Digital sound can be spread like the digital images. In audio communication, the caller and answer can conceal important information in the audio signals. The original audio can be treated as a public file, the secret can be hidden in this carrier. The secret may be text, audio or other media, generally the hidden information can be binary information. The key of information hidden is employed to embed secret and extract the secret. In the scheme, secret sharing also can be applied to hide information, an example is provided to hide digital media in audio signals below.

Low Bits Encoding   In information hiding, the bits in a byte play different roles, the LSB can be applied in information hiding. The least bits can be replaced by the bits from the secret, the key may include various transformations of the final bit sequence.

Phase Coding   In phase coding, the secret is represented by a special phase and change of the phase. If the audio signals are segmented into difference pieces, information hiding always requires:

1. The difference of two phases of two segments should be fixed, $\rho(f)$ is the phase of the i-th phase, f is the bandwidth defined on the whole wave.

2. The signal carriers should be sufficient smooth, a discontinuous sound will cause suspicious and alert.

Once the embedding procedure is completed, the last step is to embed the following corresponding phase, refresh the rest spectrum of the rest each segment, finally the relevant phase is modified. The new signals can be constructed from the new phase set. During the procedure of extraction, the hidden information may be obtained by checking the first phase. The key includes the sequence of the phase modification and the size of each segment, this encoding can be employed in digital and analogy communication.

Spread Spectrum Encoding In information hiding, robustness must be taken into consideration, therefore the hidden information is not expected to be lost because of some tempering operations such as lossy compression and cropping. In the robust approach, the most common approach used is the spread spectrum approach [21] [22]. Generally speaking, spread spectrum combines a random binary pulse $w(t)$.

In the embedding, public signals $c(t)$ combined with $a \cdot d(t) \cdot w(t)$ is adhered to $v(t)$,

$$c(t) = a \cdot d(t) \cdot w(t) + v(t) \tag{2.5}$$

where a is a factor, it is used to reduce the noise due to information hiding. The embedded binary random stream will be extracted to get the secret.

Echo Hiding Echo hiding [23] hides the secret into audio sound delay when compared to the time, the sound delay time keeps within the audible limitation. In the procedure of embedding, the sound has been segmented, the echo is embedded into each piece. In the simplest situation, the embedding can be modeled as following:

$$c(t) = v(t) + v \cdot (t - \Delta d) \tag{2.6}$$

The key is the time delay Δd and $\Delta d'$, the time difference must be restored Δd and $\Delta d'$. Extraction utilizes the coherence of two audio clips. The time delay can be obtained from the difference of the two signal peaks.

2.6   Summary

In this chapter, an introduction was given on the history, status, and trends of information hiding. Information hiding is not only based on amplitude modulation, but also based on frequency modulation and phase modulation. In information hiding, the most important considerations should be how perceivable the secret is and how robust it is against attack. Trade-offs are always expected between overall quality and how much data can be hidden successfully.

Chapter 3

Digital Watermarking

Watermarking has been and still is a very important research area within media security, it is expected to play an important role in Intelligent Property (IP) protection due to the rapid development in computer communications and the Internet. This research started in 1996 and has been intensely discussed since 2000. In this chapter, the development of digital watermarking research is discussed. Afterwards, the technologies and approaches in watermarking are introduced.

3.1   Overview

Since the 1990's, the accelerated developments in digital communication paved the way for multimedia. The commercial benefits drive the increasing trends of digital communication and networking. Intelligent Property (IP) Protection has never been as important as it is now in these times of the Internet. Digital watermarking is a common solutions to protect the ownership of digital products.

Digital products such as photographs, videos, and audio are very wide spread due to the nature of the Internet and rather easy to misuse. Fingerprinting and watermarking are recommended as the basic solutions to verify the copyright owners. Digital watermarking involves embedding a secret consisting of binary streams into a vast array of multimedia data, and use it to reserve the copyright. Watermarking is different from encryption, it keeps the integrity and recognition of the original image. Watermarking, equivalent to digital signatures, will lose its power if the host media and the adhered software disappear. Watermarks are expected to take effects in file preview, printing and transmission.

3.2 Watermarking History

SONY and Philips invented the "copyright series management device" to protect the copyright of digital audio cassettes in 1980's, it is recognized as the first device to protect copyright of digital commercial products. The aim of this product is to protect the ownership of users and encourage them to create the products of themselves.

In 1996, Adobe Systems Inc. added watermarking functionalities in Adobe Photoshop 4.0 which was developed by Digimarc Inc. At the same time, the institutes of NEC completed the software: Tiger Mark Data Blade; Informix Software finished watermarking functionalities in the database product Informix-Universal Server (Information Management System).

Europe electronic industries hoped to monitor illegally copy video and audio commercial products using watermark censoring system to find the illegal duplications. The project is called TALISMAN (Tracing Authors' Rights by Labeling Image Services and Monitoring Access Networks, it started from September 1995, 11 communication and broadcasting companies, research institutes and universities involve in it. There are two important ID adopted in the system: ID of the copyright owner which is embedded in the multimedia data; another is unique international code such as ISBN. The two IDs are expected to work jointly so as to protect the ownership.

From 1996, the International Workshop of Information Hiding was held every year, the topic of watermarking research became a hot research area [24]. In the past ten years, conferences organized by the ACM, IEEE, and IFIP, watermarking is the main research topics for media security and assurance [25]. A key point in watermark research occured when Cox et al. extended watermarking algorithms from the spatial domain, and presented a spread spectrum watermarking technique in the frequency domain [21]. Even today, this paper has been widely cited. It is thought of as the important milestone in watermarking research, it's the landmark of robust watermarking [26, 27, 28].

Audio [29, 30, 31] and video [32, 33]watermarking are also very important members in the family. In 2000, Zhao J. present a work about audio watermarking, this approach combined scrambling and DCT transformation together, it guarantees the quality of host audio while ensure the requirements of robustness [34]. The research scientists from Microsoft research Asian applied watermarking in video and presented the video watermarking based on wavelets [35].

In Siggraph' 99, Praun et al. presented robust watermarking for meshes, and introduced watermarking to computer graphics [5]. Before this work, Benedens [36], Ohbuchi [37], Yeung et al.[38] also studied this issue, Song et al. compared the usual watermarking techniques [39].

Although watermarking is designed to protect copyright by embedding secret to the host media, according to the differences of functionalities and appearance of watermarking, watermarking has been grouped into many categories. Usually watermarking is divided into visible and invisible, fragile and robust, spatial and frequency domain based.

3.2.1 Visible and Invisible Watermarking

Watermarking is grouped into two basic categories according its imperceptible to human visual system: visible and invisible. Visible watermark such as logo can be seen on the visual media such as images, photos and videos. Although invisible watermarks cannot be observed visually, invisible watermarking is used to validate or identify original authors, ownership, distributors, and even the EXIF data in a digital photo. They are all applied to protect the ownership.

Although visible watermarking reduces the commercial value of the digital products, it does not decrease usability and authentication of the media. A typical example would be television stations marking their logos at the corner of the screen while broadcasting their programmes.

It is very hard for the human visual system to detect invisible watermarking, however, the invisible watermark is extractable by computer programs. Invisible watermarking can be employed in audio, video and other digital media aside from images. For an example, watermarks can be embedded into digital audio, the owner can sensor the radio so that the radio station cannot play the illegal music disc or songs.

3.2.2   Robust and Fragile Watermarking

One of prominent attributes of digital media watermarking is to embed robust watermarks in host media. Robust watermarking refers to tampering with the watermarks in media, the watermarks can be restored from the destroyed media. In watermark design, most approaches tradeoff overall quality in exchange for strong robustness.

Tampering with a robust watermark includes following: 1) General signal processing. This includes D-A and A-D transformation, re-sampling and re-quantization, modifying the visual data in spatial domain, such as color, hue, saturation, contrast; 2) General geometry transformation, this includes: rotation, shift, cropping and zooming; 3) double watermarking (not dual watermarking). This refers to the embedding watermarks in the watermarked media again, this will cause conflictions; 4) Certain degree of generality. This means that a watermarking scheme for digital images can be employed for digital audio and video; 5) Watermarking should have clear authentication.

Similar to cryptography, there does not exist the absolutely secure cryptosystem, watermarking robustness is relative, these does not exist absolutely robust watermarking in this world, these does not exist any watermarks in this world that it cannot be destroyed.

The fragile watermark is also used for authorization and authentication. Fragile watermarking is suitable for detection of the minor changes in digital media, this is very helpful in detecting the integrity of media. The obviously destroyed media can be used to extract the fragile watermark, so as to confirm whether the media copyright has been destroyed or the media has been stolen. Fragile watermarks provide the relevant evidences in this area.

3.2.3   Spatial and Frequency Domain Watermarking

Watermarking schemes have been grouped as spatial watermarking and frequency watermarking according to the application domain. Spatial domain watermarking usually refers to embed watermarks in pixels of visual products in various color space, sometimes watermarks can be embedded into luminance, saturation and contrast. A very important watermarking form is LSB [40, 41]. Schyndel's paper published in ICIP'94 is thought as the world first paper about watermarking. A lot of research scientists have also suggested to select pixels randomly, so as to embed watermarks in pixel blocks [42, 43].

On frequency domain applications, current technologies are still focused on DCT and DWT transformation. These transformation embedded watermarks in the coefficients of frequency transformations, the media is re-constructed by using inverse

transformation. The watermarks are normally extracted from coefficients of the attacked media, the watermarks are identified by using statistics. Frequently asked questions are: whether any watermark exists here? What is it [44]?

3.2.4 Watermarking Approaches

Mintzer invented the basic LSB approaches: Suppose there is a 24-bit (3×8bytes) image, a watermark is binary stream, the least significant bit is replaced using one bit of the watermark, this approach adopts the principle that the changes of least significant bits do not impact the visual quality of the images, it is regarded as the basic law of digital watermarking.

The frequency approaches based on Cox's approach are thought as the main stream of robust watermarking in frequency domain, the general steps of this approach are:

Suppose file D has a vector $V = (v_1, v_2, \ldots, v_n)$, the watermark $X = (x_1, x_2, \ldots, x_n)$ is embedded in it and a vector, $V' = (v'_1, v'_2, \ldots, v'_n)$ is obtained, the new V' is used to replace V in the document to get the watermarked file D'. D' may suffer any kind of attack so that it is only possible to get the tampered file D^*, if the original D is known, and the attacked file D^* is also known, the tampered watermark X^* can be extracted, and compare it with the original watermark X, so as to confirm whether the watermark has survived the corresponding attack.

Embedding Procedure

Watermarking usually adopts one of following formula:

$$v'_i = v_i + \alpha x_i \tag{3.1}$$

$$v'_i = v_i(1 + \alpha x_i) \tag{3.2}$$

$$v'_i = v_i(e^{\alpha x_i}) \tag{3.3}$$

The equations are employed in different ways, Eq. 3.2 and Eq. 3.3 are suitable for the status where v_i is very small, Eq. 3.1 can deal with the situation where v_i is greater.

Extraction Procedure

The extraction procedure obtains the vector $V^* = (v_i^*, v_2^*, \ldots, v_n^*)$ from the document D^* using frequency transformation. Then $V^* = (v_i^*, v_2^*, \ldots, v_n^*)$ and $V = (v_1, v_2, \ldots, v_n)$ are calculated to get the modified watermark X^*.

Detection Procedure

The detection procedure is based on statistics, the similarity of two watermarks is defined as follows:

$$sim(X, X^*) = \frac{X^* \cdot X}{\sqrt{X^* \cdot X^*}} \tag{3.4}$$

Finally the threshold T is given by statistics, if $sim(X, X^*) > T$, the watermark exists, otherwise, it does not.

3.3 Video Watermarking

In video watermarking, Hartung et al presented the watermarking based on MPEG-2 [45, 46, 47], Swanson et al. developed watermarking approaches MPEG-2 and wavelets for multiple resolution [48]. Although the details are diversity, however the core is orthogonal transformation, such as DCT and DWT. These transformations convert various color spaces to frequency domain, the watermarks are embedded into the coefficients. These coefficients are stable in frequency domain. With the development of video compression technologies, MPEG-4 and H.264 are employed to host watermarks [49].

Cox watermarking comes from spread spectrum, namely watermarks and noise are modulated together and added into the host media. In MPEG files, watermarks can be added into the DCT coefficients, this allows watermarks and video frames to be encoded and decoded together. To void the visual quality losing and making sure the watermark invisible, watermarks can be treated like the general signals and even can be resisted attacks such as Stirmark.

Wavelets have important usage in JPEG2000. Usually in watermarking, watermarks are embedded into HH coefficients of the wavelets, the media has not obvious changes. Because wavelets transformation generates lots of coefficients in multiple resolution, the volume of watermarks and volume of media data can be very large. Evidences show that watermarks can be correctly identified in any resolutions.

In practical operations, suppose the resolution of image is $N \times N$, $R + 1$ is the wavelet layers, watermark vector is: $X_N = \{x_1, \ldots, x_{n1}, \ldots, x_{n2}, \ldots, x_{nR}\}$. Under different resolutions, it has the following relationships $X_1 \subset X_2 \subset \cdots \subset X_R$. In order to embed watermark X into the high pass band, and combine with the V together, $v'_i = v_i(1 + \alpha_i x_i)$. Zooming factors α_i can change the energy intense, it takes important role in invisible and robust. In order to detect watermark X^i from V', the similarity between X' and X can be determined, $sim(X, X') = X' \cdot X / \|X'\|_2$. If the watermarks match each other, a further threshold will be provided t. $sim(X, X')$ is subject to the standard norm distribution.

3.4   Video Logo Erasing

A video logo, is a trademark or a symbol of Intellectual Property Protection (IPP) and is popularly used by most television channels. It helps in identifying the owner of the video clips. Figure 3.1(a) gives some examples of video logos. Generally speaking, a video logo possesses the following characteristics:

- A video logo has a special shape that makes it different from other video logos.

- A video logo should be visible. It should be distinguishable from the background color.

- A video logo has a stable position in video frames, it does not move with time and events.

- A video logo area should not be large and should not affect the visual effects of video content.

- A video logo is not altered for a long time.

However, a video logo sometimes results in annoying visual artifacts especially in cases where multiple logos exist in filed videos that are reused and in news videos obtained from some other channel. Multiple logos in video clips affect the visual pleasure like in cases where some television stations superimpose their video logos on other logos or when others mark the logo area in video frames and overlap it with their own logos. In Figure 3.1(b), the logos are overlapped directly, stacked on another mosaic style logo, and blended with the blurred one. This is visually not very appealing to the viewers. These artifacts provided the motivation for the research and the aim is to erase video logos from video frames. This is also very useful as an attack on visible watermarks in digital videos.

(a) Video Logo Examples.

(b) Video Logos Superimposed.

Figure 3.1: Sample Video Logos.

After the video logo area is marked the logos can be erased from the video frames based on image inpainting. This problem in [6] and [50] is described as a discrete approximation of the partial differential equation (PDE):

$$I_t = \bigtriangledown^{\perp} I \cdot \bigtriangledown \triangle I \qquad (3.5)$$

where $\bigtriangledown^{\perp}$ denotes the perpendicular gradient $(-\partial_y, \partial_x)$ and $\triangle$ denotes the Laplacian operator $\partial_{xx} + \partial_{yy}, I$ is the image intensity, $I_t = \frac{\partial I}{\partial t}, \frac{\partial}{\partial t}$ is $\frac{\partial}{\partial t} + v\bigtriangledown, v = \bigtriangledown^{\perp} \triangle I$. Equation 3.5 projects the gradient of the smoothness of the image intensity in the direction of the isophote [50]. In order to solve the PDE, compute-intensive algorithms are provided in [6] and [50] based on fluid dynamics.

Figure 3.2 is helpful in understanding the inpainting algorithm. In the Figure 3.2(a), every circle is marked with an alphabetic character indicating a pixel in the video frame. The region with "B" indicates that the area needs to be inpainted. The

(a) Complete Inpainted Area.   (b) Partial Inpainted Area.
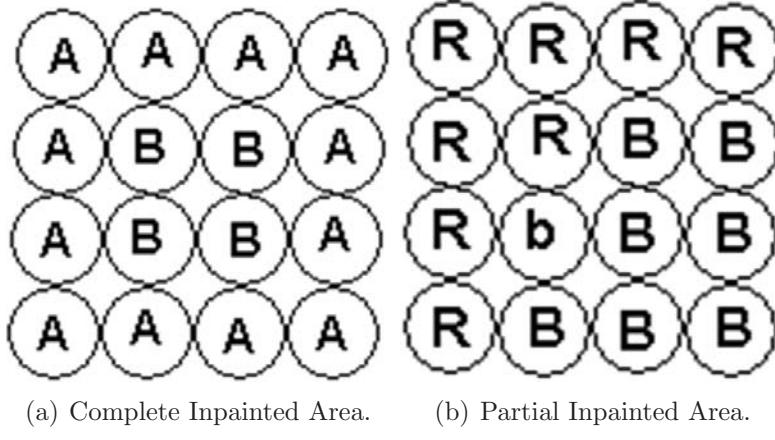
Figure 3.2: Description of the Image Inpainting Algorithm.

region with "B" and "b" in Figure 3.2(b) has to be filled with the information from the surrounding pixels with "R". The region with the "R" pixels represents the undamaged area of an image. When it is decided that the "b" pixel should be filled, the colors are averaged of the surrounding "R" pixels. For example in Figure 3.2(b), the "b" pixel is surrounded by "R" pixels in 8 directions. The "b" pixels are filled with the average color of all these surrounding "R" pixels.

Suppose only one pixel needs to be filled, the average color of the surrounding pixels represents the filled color. If there are many rings, all pixels from the outside ring to the inside ring are filled by the average color. The remaining unfilled pixels, if any, are filled by the average colors of the surrounding pixels. These are filled according to the scanning order from left to right and top to bottom. After ring filling and scanning filling are over, the region is completely filled.

The average color can retain a better visual effect. Assume the color of a grey pixel to be filled is $I(i, j)$, then,

$$I(i, j) = \frac{1}{N} \cdot \sum_{k=1}^{N} I_k(i, j) \tag{3.6}$$

where $I_k(i, j), i = 0, 1, \ldots, R_x - 1, j = 0, 1, \ldots, R_y - 1$ is the grey color of surrounding pixels about $I(i, j)$ not to be pixels. So the error of filled color to the color of the filled, $N \leq 8$ is the total number of these surrounding pixels is:

$$\varepsilon = \sum_{k-1}^{N} |I(i, j) - I_k(i, j)| \tag{3.7}$$

$$\varepsilon \le N|I(i,j)| + \sum_{k=1}^{N} |I_k(i,j)| \le 2\sum_{k=1}^{N} |I_k(i,j)| \qquad (3.8)$$

Eq. 3.7 and Eq. 3.8 tell us that the error is related with the color of surrounding pixels, if the surrounding color has little variance, the error is little and the area can be filled more accurately.

(a) Image with logo to be removed.   (b) Image with corresponding logo removed.

Figure 3.3: Logo erasing.

3.5   Logo Removal using Video Inpainting

Image inpainting technique could be considered for logo removal. However, since this method cannot erase logos that have a large area, a different approach is examined in this section. The video logo is removed by obtaining the most similar region in the video frame to fill the region occupied by the logo. It is akin to the idea of motion compensation used in compression albeit with an idea of doing region restoration. Although many distance measures can be selected for matching, herein the matching method based on the Hausdorff distance is used for the sake of simplicity. The advantage of this proposed system is that it is automatic without human intervention. Moreover, we define a new concept: video inpainting. In this approach, we imagine the logo in a frame sequence to be a cylinder and the job is to fill the cylinder with the help of neighboring frames and boundary information.

3.5.1   Matching Based Logo Removal

Given two finite point sets $A = \{a_1, a_2, \ldots, a_m\}$ and $B = \{b_1, b_2, \ldots, b_m\}$, the Hausdorff distance is defined as: $h(A, B) = \max(h(A, B), h(B, A))$, $h(A, B) = \max_{a \in A} \min_{b \in B} ||a_i - b_j||$, $a_i \in A$, $b_j \in B$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$ and where $|| \cdot ||$ is a norm on the points of A and B.

The key advantage of using the Hausdorff distance to compare the difference of images is its low sensitivity to small perturbations of the image. The computation of Hausdorff distance differs from many other shape comparison based methods in which there is no correspondence between the model and the image. The method is quite tolerant of small position errors such as those that occur with edge detectors and other feature extraction methods [51, 52, 53, 54]. Although the pixel position information is not used in the computation of Hausdorff distance, each pixel has its fixed position in the blocks related to logo searching. The position of each search block is the primary concern instead of the position of each pixel.

Assume that a video logo occupies a portion of video frames $I_w(x, y)$ in image $I(x, y)$, the most similar region $I_s(x, y)$ should be found in the image $I(x, y)$ among all video frames under the definition of Hausdorff distance. The content of this region is the most similar to the content of the region that the logo occupies $I_o(x, y)$. Namely the similarity takes the maximum value in all the candidate regions $I_c(x, y)$:

$$d_c = \frac{I_o \cdot I_c}{|I_o||I_c|}, d_s = \min(d_c), I_s = I_c|_{d_c = d_s} \tag{3.9}$$

If this region is found, it can be used to replace the logo. This method has a problem since the logo overlaps the region, finding the best match for it is problematic. Since

it is not known what lies underneath the logo, adjacent regions are considered. The regions to the right, left, top, and bottom of the video logo region are considered for matching with other frames in the shot. For the current frame, these four regions are matched with all the other video frames. The most similar pairs are found among all the candidate pairs in the video frames. The corresponding position of video logo in the best-matched frame will be found. For instance, if the highest similar pair among the candidate regions is on the top of logo region in a video, the region under the highest similar region of the matched frame will replace the logo region. Subsequently the region with the logo will be replaced and thus the video logo is removed.

When this algorithm is utilized to remove video logo, at least three frames are considered; they are the current frame, next frame and previous frame in the same shot if available.
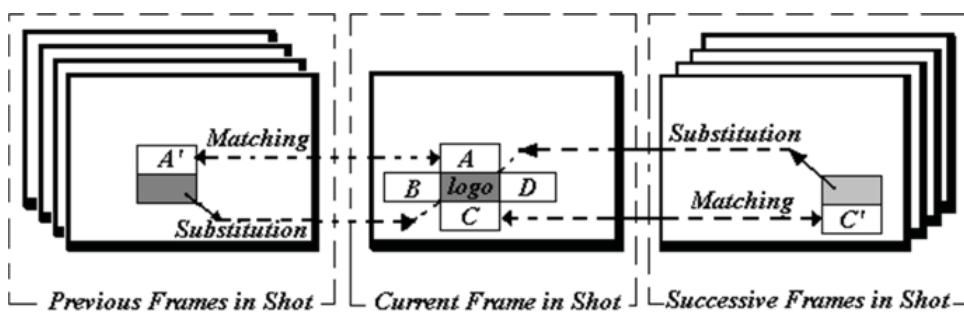
Figure 3.4: Matching based algorithm for region overlapping.

In Figure 3.4, suppose the middle frame is a frame of the input video with a logo. Around the video logo, four regions are marked with **A**, **B**, **C** and **D**. In order to obtain the substitution region of the logo, the regions in the related frames (previous frames and next frames) of the current frame are considered with the most similar pair among all the candidate pairs being selected. The corresponding region is then used to replace the watermarked region. If the most similar pair is on the top of the logo, the region under matching is utilized to overlap the visible watermark, shown in the left part of Figure 3.4; if the most similar pair is at the bottom of the logo, the area above the matched region is utilized to overlap the logo as shown in the right part of Figure 3.4.

Since the objects in video are moving, the region occupied by the video logo in the current frame may not be obscured by the logo forever. Thus the opportunity is there to locate that region without a logo covering which can be subsequently used for removing the logo from the current frame.

The drawback of this technique is similar to that of locating the logo region. If most objects in the frames of target video are stationary, perhaps a better substitution

region cannot be found to replace the region occupied by the video logo. However it is possible to find some regions to replace the logo even if it is not the optimal one. Another shortcoming of the proposal is that sometimes the found blocks will bring extra visible edges after overlapping and destroy the coherence of this region. Therefore, the video frames should be inpainted using video inpainting. In short, overlapping based logo removal is a supplementary measure to handle the logos with motion in the background (for at least one logo region position). If the motion is not so significant, video inpainting can be employed to fill the logo region.

3.5.2 Video Inpainting Based Logo Removal

Unlike image inpainting and block based overlapping algorithms for video logo removal, video inpainting considers a frame sequence to fill a logo region. The logo region to be filled in the frame sequence is treated as a cylinder. The processing of video logo removal can be considered as the filling of cylinder in the video volume layer from the outmost to the innermost layers. The scheme is illustrated by Figure 3.5. Three example frames of video are shown in the diagram. The rectangles represent pixels in frames with the red ones indicating the current pixels to be inpainted. In Figure 3.5(b), the green rectangles illustrate the pixels those are needed to be inpainted later and the information of the blue pixels is utilized to inpaint the current pixels. The yellow rectangles are also utilized in the rendering of the current pixels.

(a) Cylinder layer filling based video logo removal. (b) Each pixel on the layer of cylinder is inpainted by using adjacent layers. (c) Direction of the axis towards in the cylinder.
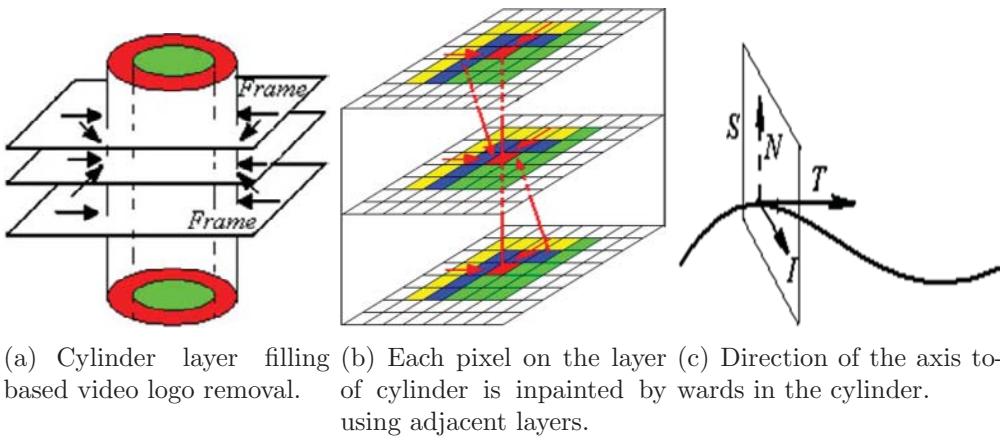
Figure 3.5: Video based inpainting logo removal.

In Figure 3.5, the logo region is overlapped by using a 3-dimensional gradient vectors instead of 2-dimensional gradient vectors normally used in image inpainting to obtain a region to replace the logo. The contributions of the gradient vectors are computed based on adjacent frames in order to predict the color value of the current pixels.

Motion vectors are also used for logo regions in order to find the pixels on the outside rings. The idea is to fill the marked region more precisely than those proposed before.

Mathematically, suppose Ω is the logo region with $\partial\Omega$ being its boundary. A pixel on the boundary $\delta\Omega$ that needs to be inpainted is denoted as $y = f(x, y, t)$, where (x, y) is the pixel position, t is the frame on which it is located. The tangent vector is $T(f_x, f_y, f_t)$ and the normal vector $N(n_x, n_y, n_t)$ lies on a plane, where $T \cdot N = 0$. The gradient vector $G(-n_x, -n_y, -n_t)$ also is on this plane. In order to inpaint the logo cylinder, only those normal vectors $I(I_x, I_y, I_t)$ are utilized, whose direction points towards the axis of the cylinder, namely: $G \times (0, 0, 1) \cdot T > 0$ as shown in Figure 3.5(c). The inpainted pixels are estimated by using the first order difference:

$$\triangle f_t = f(x, y, t + 1) - f(x, y, t) \tag{3.10}$$

$$\triangle f_x = f(x + 1, y, t) - f(x, y, t) \tag{3.11}$$

$$\triangle f_y = f(x, y + 1, t) - f(x, y, t) \tag{3.12}$$

$$
\begin{aligned}
f(x, y, t) = \quad & w_t \cdot (2 \cdot f(x, y, t - 1) - f(x, y, t - 2)) + \\
& w_x \cdot (2 \cdot f(x - 1, y, t) - f(x - 2, y, t)) + \\
& w_y \cdot (2 \cdot f(x, y - 1, t) - f(x, y - 2, t))
\end{aligned} \tag{3.13}
$$

where the $w_t \in [0, 1]$, $w_x \in [0, 1]$ and $w_y \in [0, 1]$ are weights that indicate the contribution of the various gradients to the current pixels, and $w_t + w_x + w_y = 1$. In temporal domain, video frame coherence is exploited and the correlation between these frames is utilized. Thus the 3D gradients from different directions must be computed. Not only are the 2D gradients used to compute the influence on the current pixels; the temporal domain information from adjacent frames is also used.

The video inpainting based method fills the concentric cylinder step by step along both the temporal and spatial directions by propagating the region information. The video inpainting processing can be visualized as the dissolving surface of a cylindrical "ice block" in a tank of water. One of challenges of this 3D dissolution method is that it is needed to obtain the motion estimation; this is used to infer the optical flows and isohyets.

3.6 Summary

Digital watermarking has been regarded as an important tool to protect the IP ownership. In this chapter, the history of digital watermarking was introduced. The research in the area of digital watermarking is very helpful to protect the copyright of digital products in this Internet age.

Chapter 4

Digital Scrambling

Scrambling a digital signal in the spatial or frequency domain corresponds to altering that signal in such a way that the original semantic media loses its meaning and becomes difficult to view. This technique has been successfully applied in the television industry. This chapter starts at looking at digital image scrambling, then progressive audio scrambling is introduced. The approaches presented in this chapter can be applied in many situations. Descrambling a digital signal is very easy to implement, it is usually the inverse procedure of the original scrambling function.

4.1   Overview

Scrambling refers to transforming a semantic piece of media data into something that contains sufficient disorder to make the original semantic meanings disappear. Its inverse transformation is called descrambling. The attackers cannot get the original media even if they have the scrambled media. Scrambling should satisfy (1) low degree of understanding; (2) high quality; (3) no change in the frequency; (4) difficult to be broken.

4.2   Digital Image Scrambling

Digital image scrambling originated from scrambling of cable TV signals. With the development of digital high definition TV, scrambling has attracted much more attention. This is due to the fact that scrambling is an important means of digital media security, in that it can transmit a scrambled signal and restore it successfully with no quality loss.

Table 4.1: The chessboard order of 26 letters.

A	B	C	D	E
F	G	H	IJ	K
L	M	N	O	P
Q	R	S	T	U
V	W	X	Y	Z

Digital scrambling refers to transforming an image into another completely different image, the operator only knows the algorithm and keys, this allows them to restore the original image.

If any user does not have any prior knowledge of the algorithm and the corresponding keys, the original image cannot be restored. Image scrambling looks very easy, in fact it is really not like this. This is because the algorithms are not secure and cannot ensure the scrambled images lose the original semantic meanings. Digital image scrambling requires that the scrambled image have the same size as that of the original image. Scrambling does not affect the image resolution.

Compared to the research objectives of cryptography, a digital image provides a large amount of data, or in other words, it can provide a larger space for plain text and cipher text. The most important thing is that the self-correlation of the digital image is represented in two directions perpendicular with each other, while the self-correlation of a text, such as one-dimensional signal sequence, is rather difficult to obtain. With these two characteristics of the digital image, attackers may try to find a fixed frequency of each pixel, but, since different images have different histograms, it will be very difficult for them to find it. How can a map be established of the conversion between the original image and disguised one? In digital scrambling, the attackers wish to find out the frequency of each pixel, since different images have different histograms, therefore, such a search is very hard.

In [55], the past and current encryption schemes have been summarized, such as random scan line scrambling, scan line transposition scrambling, scan line cycling scrambling, and pixel scrambling. Matias and Shamir presented a scrambling technology based on curve filling-up [56].

4.2.1   Chessboard Coding Based Scrambling

In 2000 B.C., the Greeks created chessboard coding. This ancient code has been prevalent around the world. In the code, 26 letters are put in 5×5 grids, where i and j are put in the same table cell in Table 4.1:

Thus, each letter corresponds to a group of numbers consisting of the row and

column numbers, e.g. "m" corresponds to "32". Therefore the following text can be expressed as:

Text: Digital Image Security Processing

Code: 14 24 22 24 44 11 31 24 32 11 22 15 43 15 13 45 42 24 44 54 35 42 34 13 15 43 43 24 33 22

Corresponding to the 256 color values of a grayscale digital image, there exists such a table that maps one color value onto another. Therefore each color has a number, the table is transposed and another image is obtained. Figure 4.1 is an example using this chessboard based image scrambling.
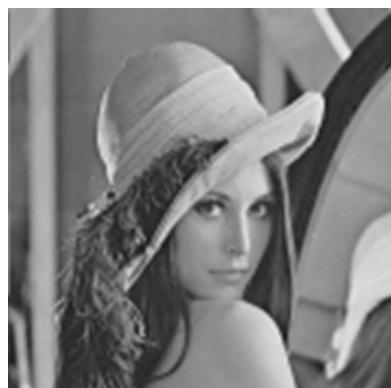
(a) Original image.   (b) Scrambled image.

Figure 4.1: Image scrambling based on chessboard scrambling.

The chessboard coding scrambling scheme can be generalized which allows the problem to be described as follows:

Given matrix A and image I, image I should be re-ordered according to matrix A within the color domain, spatial domain or frequency domain, the scrambled image is completely disordered, any observers cannot acquire any semantic information from it. Figure 4.2 is an example of using this chessboard coding based scrambling, the matrix in this example is generated using Hilbert curves.

(a) Letter "A".   (b) H-curve.   (c)   Scrambled
                                                           "A".
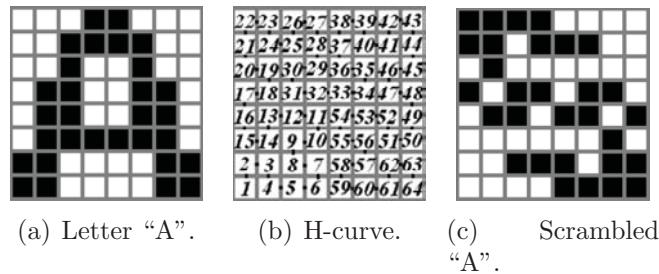
Figure 4.2: Image scrambling based on chessboard scrambling.

4.2.2   Caesar Coding Based Scrambling

In cryptography, Caesar coding is defined as follows:

$$c = (k_1 m + k_2) \bmod 26 \tag{4.1}$$

Where m is text, c is cipher text, k_1 and k_2 are keys. The modulus is 26, as there are 26 letters in the English alphabet.

Eq. 4.1 can be extended to a more general situation, the following formula is obtained:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \left(\begin{bmatrix} R_1 \\ G_2 \\ B_3 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \right) \bmod 255 \tag{4.2}$$

where $\beta_1, \beta_2, \beta_3$ are different parameters, they are called keys, `mod` is the modulus operation, (R, G, B) is the scrambled color, (R_1, G_2, B_3) is the original color. Figure 4.3 and Figure 4.4 are the scrambling results using Eq. 4.2.

This approach can be generalized for image scrambling based on the Arnold transformation, the Arnold transformation is:

(a) Original image.   (b) Scrambled image.
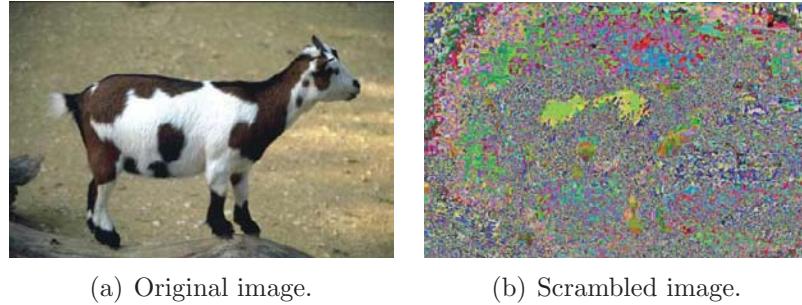
Figure 4.3: Example of scrambling.

(a) Original image.   (b) Scrambled image.
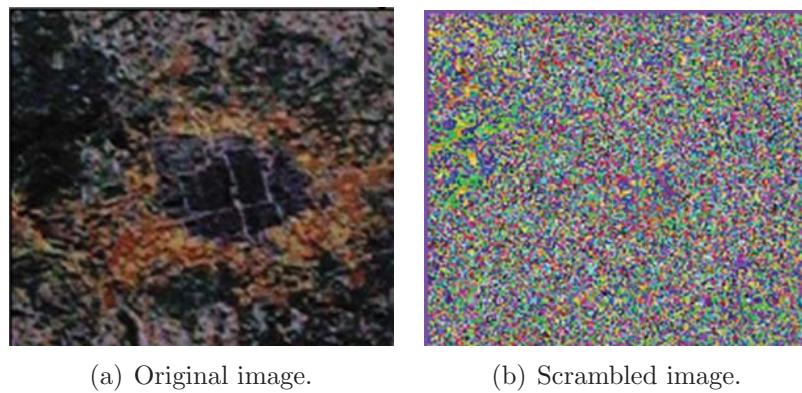
Figure 4.4: Image scrambling.

$$C = TM \bmod n \qquad (4.3)$$

Where M and C are n-dimension vector. T is $n \times n$ matrix, n is integer.

$$T = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & 2 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 2 & \ldots & n \end{bmatrix}_{n \times n} \qquad (4.4)$$

where $n = 2$ and $n = 3$, T are respectively:

$$T = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}_{2 \times 2}, T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}_{3 \times 3} \qquad (4.5)$$

Figure 4.5 is the result of synthetically applying Eq. 4.5 in position scrambling and color scrambling.
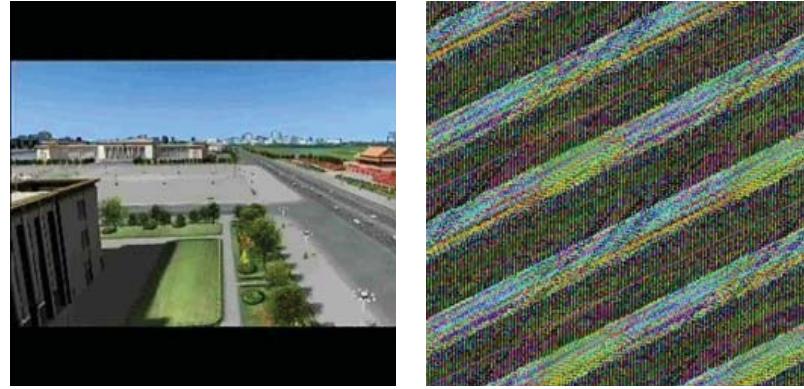
(a) Original image.   (b) Scrambled image.

Figure 4.5: Synthetic scrambling based on Arnold transformation.

4.2.3   DES Based Scrambling

XOR operation refers to:

$$1 \oplus 1 = (1+1) \bmod 2 = 0; 1 \oplus 0 = (1+0) \bmod 2 = 1;$$
$$0 \oplus 1 = (0+1) \bmod 2 = 1; 0 \oplus 0 = (0+0) \bmod 2 = 0;$$

where $\oplus$ is noted as XOR operation. The advantage of using XOR operation is repeating XOR operations will restore the orignial number. e.g. $110010 \oplus 101101 = 11111; 11111 \oplus 101101 = 110010$.

Suppose an image is taken and a key is selected to generate a random image. Using the XOR operation to combine both images on the color space, the result will be indistinguishable. If the scrambled image is XOR'd again with the random image, the original image can be recovered. Figure 4.6(c) is the scrambled image using XOR operation between Figure 4.6(b) and 4.6(a). The advantage of this scrambling algorithm is that it is swift, convenient, the key space is very small, the security of the scrambled image has to be subject to the key and the relevant parameters.

Single direction scrambling based on XOR operations cannot generate very good scrambling quality, the reason is the outline of the image still visible. Therefore, the scrambling based on double direction XOR operations is shown in Figure 4.7(b). The scrambling algorithm based on double direction XOR operation sufficiently scrambles the original image, the scrambled image successfully hides the outline of the image. This guarantees the scrambled image is indiscernible in the color space.

Figure 4.8(a) is the original image, Figure 4.8(b) is the scrambled result using single direction XOR operation based scrambling, the outlines can be clearly seen. Figure

(a) Original image.   (b)   Randomly   generated   (c) Scrambled version of the
                           image.   original.
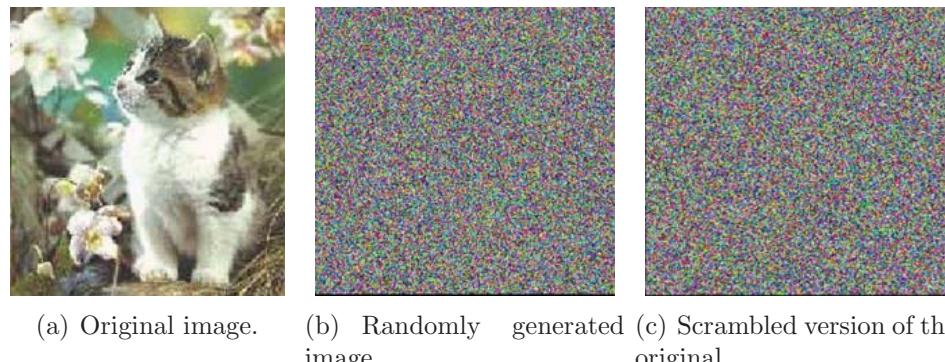
Figure 4.6: XOR operation based image scrambling.

(a) Single direction scram-   (b) Double direction scrambling.
bling.
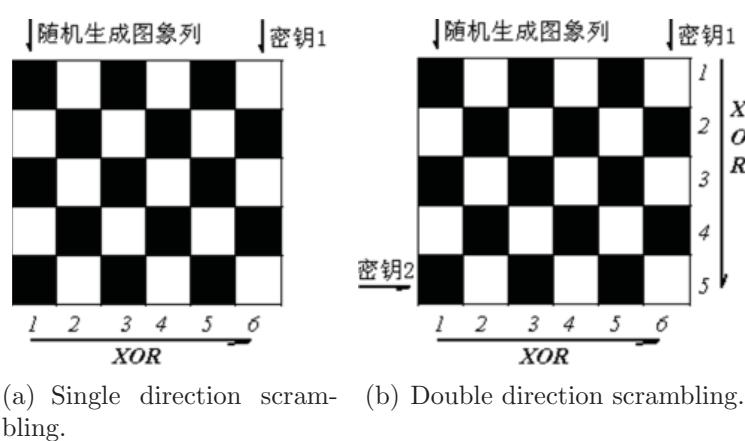
Figure 4.7: Improved image scrambling based on XOR operations.

4.8(c) is the scrambled result using the double direction XOR operation based scram-
bling. From the scrambled image, the outline cannot be found from the scrambled
image.

(a) Original image. (b) Single direction scrambling. (c) Double direction scrambling.
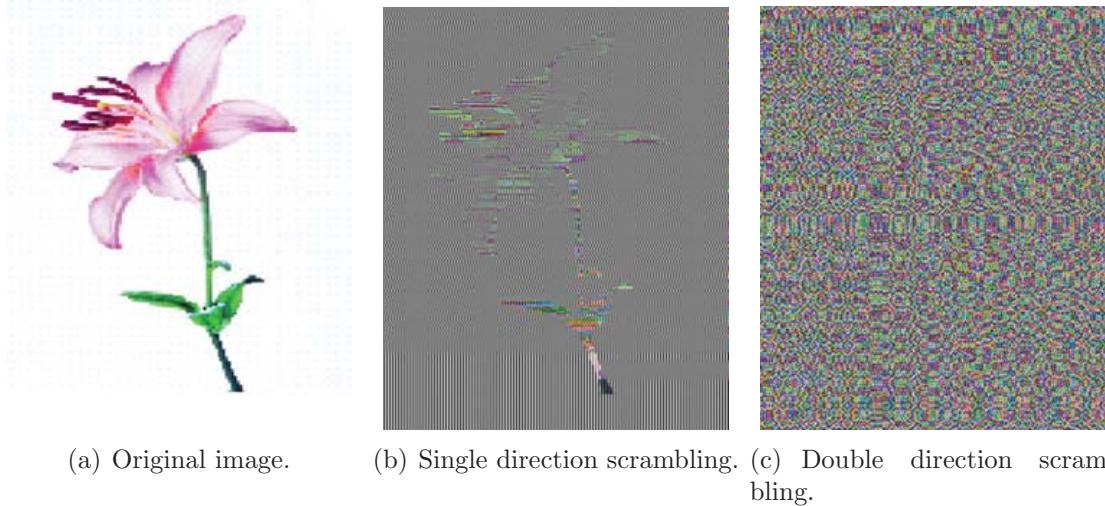
Figure 4.8: Image scrambling after XOR operations.

4.2.4 Digital Image Scrambling Based on Magic Squares

Magic square is an old mathematical problem, of which record can be found in old books. It has many useful properties and fantastic structure that it has attracted the attention of numerous scholars. The standard magic square is an n order matrix with natural numbers $1, 2, \ldots, n^2$ as its elements:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \tag{4.6}$$

which satisfies the following equations:

$$\sum_{j=1}^{n} a_{ij} = c \ (i = 1, 2, \ldots, n) \qquad \sum_{i=1}^{n} a_{ij} = c \ (j = 1, 2, \ldots, n)$$

$$\sum_{i=1}^{n} a_{ii} = c \qquad \sum_{i+j=n+1} a_{ij} = c \tag{4.7}$$

where

$$c = \frac{n^2(n^2 + 1)}{2}$$

of which the matrix A is referred to as the standard magic square.

Suppose the n-order matrix of a given digital image corresponds to B. For a fixed n-order magic square A, a map between B and A is constructed. A sequence of operations is established in A: moving element 1 to the position of element 2, move element 2 to position of element 3. Generally, for each element $m \in \{1, 2, \ldots, n^2 - 1\}$ in A, when it is moved from its position in A into the position of element $m + 1$, if $m = n^2$, then it can be moved to the position of element 1. After this kind of moving, the magic square A changes to a matrix A_1, denoting as $A_1 = EA$; for A_1, $A_2 = EA_1$ is obtained by repeating the above procedure, and then, iteratively, $A_3 = EA_2$, $A_4 = EA_3$, etc. This is known as a sequence of scrambling transformations. As a result, $A_{n^2} = A$ will be obtained after n^2 steps of iteration.

For a digital image matrix B, taking into account the correspondence of elements between B and A, elements (grayscale values of each pixel) of B are moved to corresponding position to obtain a new digital image matrix B_1, denotes as $EB = B_1$, while transforming A to $A1$. Similarly, $B_m = E^m B$ is possible. It has been proved that the cycle of this kind of digital image scrambling is as follows:

$$T = N^2 \tag{4.8}$$

The experimental result of image scrambling based on magic square is shown in Figure 4.9.

(a) Original image. (b) Iterated step 1. (c) Iterated step 2. (d) Iterated step 3. (e) Iterated step 4.
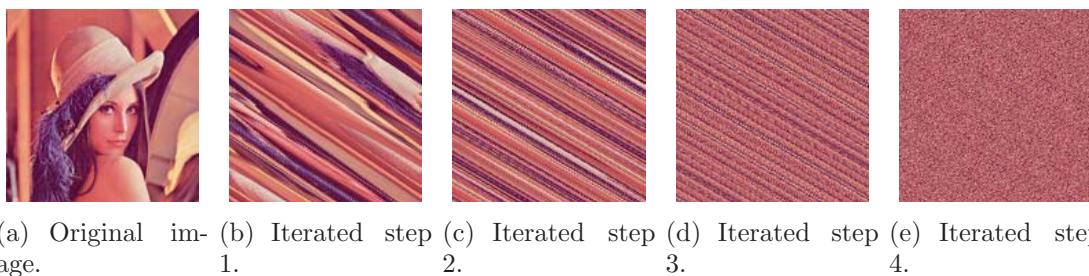
Figure 4.9: Digital image scrambling based on magic squares.

4.2.5   Digital Image Scrambling Based on Gray Code Transformation

Gray code is an integral mapping, and Gray code transformation is a type of binary representation of digital data, which can be used for error correction and verification of binary data. This section will discuss how to give the definition of Gray code

Table 4.2: Order of pixels and Gray transform table.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$(n)_2$	0	1	10	11	10 0	10 1	11 0	11 1	10 00	10 01	10 10	1011 11	11 00	11 01	11 10	11 11
$G(n)_2$	0	1	11	10	11 0	11 1	10 1	10 0	11 00	11 01	11 11	11 10	10 10	10 11	10 01	10 00
$G(n)$	0	1	3	2	6	7	5	4	12	13	15	14	10	11	9	8

transformation in matrix format, how to generalize it, and finally, how to use Gray code transformation to scramble digital images.

For any given non-negative integer n, its binary format can be written as $n = (n_p n_{p-1} \dots n_1 n_0)$, where $n_j \in 0, 1$. From such an addition: $0 \oplus 1 = 1 \oplus 0 = 1, 0 \oplus 0 = 1 \oplus 1 = 0$ the following transformation is given:

$$g_j = n_{j+1} \oplus n_j, \ j = 0, 1, \dots, p \ g = (g_p g_{p-1} \dots g_1 g_0)_2 \qquad (4.9)$$

thus obtaining $g = G(n)$, G is called the Gray transformation, and g is known as the relative Gray code of n. In Table 4.2, the result of Gray transformation from 0 to 15 are listed.

Gray transformation has a hierarchy of fine stratification, thus assuming a structure of self-similarities. In fact, Gray transformation can be expressed as following matrix format:

Theorem 4.2.1.

$$I_1 = \widetilde{I}_1 = 1, I_{2^k} = \begin{pmatrix} I_{2^{k-1}} & \\ & \widetilde{I}_{2^{k-1}} \end{pmatrix}$$

$$\widetilde{I}_{2^k} = \begin{pmatrix} & I_{2^{k-1}} \\ \widetilde{I}_{2^{k-1}} & \end{pmatrix} k = 1, 2, \ldots \tag{4.10}$$

where matrix I_{2^k} and $\widetilde{I}_{2^k}$ are block diagonal matrix and anti-diagonal matrix, respectively, and each sub-block matrix I_{2^k} and $\widetilde{I}_{2^k}$ can be divided into a half order diagonal matrix and anti-diagonal matrix.

As shown in the Figure 4.10, 4.10(a) is $g = G(n)$ represented in the rectangle coordinates system, 4.10(b) is the matrix I_{32}, where the black stands for 1, and the white stands for 0, and the solid line represents the hierarchy of fine stratification and subdivision structures. The curve in Figure 4.10(c) is obtained from the matrix of Figure 4.10(b), having the self-similarity.
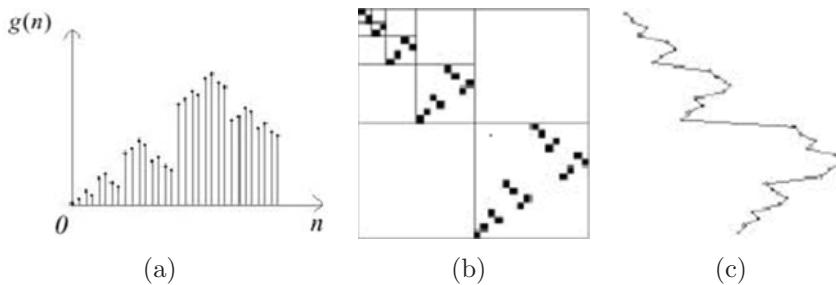
Figure 4.10: Self-similarity structure of Gray transformation.

Theorem 4.2.2. *Suppose the Gray code of nonnegative integer n is $G(n)$, let $G^0(n) = n, G^k(n) = G \circ G^{k-1}(n), k = 0, 1, 2, \ldots$. Then, when positive integer n, it satisfies $2^{k-1} \leq n \leq 2^k - 1, G^{2^k}(n) = n$.*

From Theorem 4.2.2 it is known that, for any non-negative integer n, if it can satisfy $n \leq 2^k - 1$, then $G^{2^k}(n) = n$. Thus, for any image with a grayscale level of 2^k, if a continuous transformation process is performed for its gray level, an accurate image can be reproduced.

Gray transformation can be expanded to a series of other transformations, which can be represented in the following matrix forms:

$$E_{2^k} = \begin{pmatrix} & E_{2^{k-1}} \\ \widetilde{E}_{2^{k-1}} & \end{pmatrix}, \widetilde{E}_{2^k} = \begin{pmatrix} E_{2^{k-1}} & \\ & \widetilde{E}_{2^{k-1}} \end{pmatrix}, k = 1, 2, \ldots \qquad (4.11)$$

$$F_{2^k} = \begin{pmatrix} F_{2^{k-1}} & \\ & \widetilde{F}_{2^{k-1}} \end{pmatrix}, \widetilde{F}_{2^k} = \begin{pmatrix} & \widetilde{F}_{2^{k-1}} \\ F_{2^{k-1}} & \end{pmatrix}, k = 1, 2, \ldots \qquad (4.12)$$

$$G_{2^k} = \begin{pmatrix} & G_{2^{k-1}} \\ \widetilde{G}_{2^{k-1}} & \end{pmatrix}, \widetilde{G}_{2^k} = \begin{pmatrix} \widetilde{G}_{2^{k-1}} & \\ & G_{2^{k-1}} \end{pmatrix}, k = 1, 2, \ldots \qquad (4.13)$$

The other transformations can be obtained by commuting the position of $I_{2^{k-1}}(E_{2^{k-1}}, F_{2^{k-1}}, G_{2^{k-1}})$ and $\widetilde{I}_{2^{k-1}}(\widetilde{E}_{2^{k-1}}, \widetilde{F}_{2^{k-1}}, \widetilde{G}_{2^{k-1}})$ in expressions 4.10-4.13.

In the above mentioned Gray transformation matrix, it is natural for us to present a kind of digital image scrambling method, which can act not only on the position space of the digital image, but also on the color space or frequency space of the image. The following Figure 4.11 gives the result of the image scrambling using the above equations 4.10-4.13.

(a) Scrambling by Eq. 4.10   (b) Scrambling by Eq. 4.11   (c) Scrambling by Eq. 4.12   (d) Scrambling by Eq. 4.13
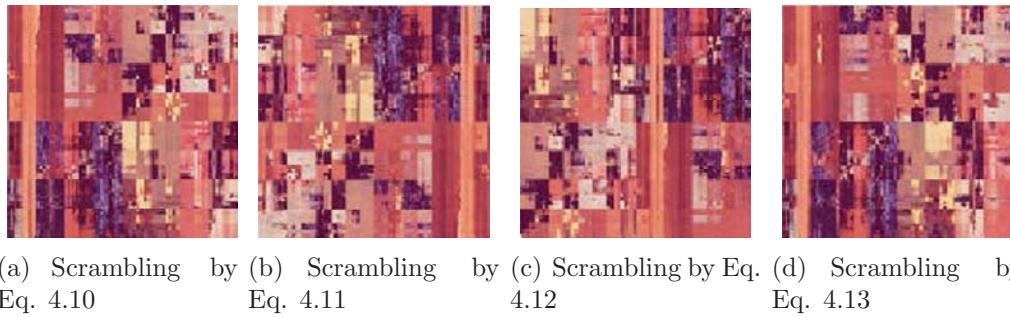
Figure 4.11: Digital image scrambling based on generalized gray transformation.

4.2.6   Digital Image Scrambling Based on Conway's Game

Conway's game is a special image matrix transformation. In the 1970's, British mathematician John Conway and his students confirmed apt rules after repetitious experiments. Suppose there is a planar grid, in which each node represents a life cell. If it is filled with black color, the cell is believed to be alive; if it is filled with white color, the cell is considered dead. Suppose the planar grid is infinite great (in fact, the upper and bottom, left and right boundaries of a finite planar grid can be linked up), then each cell will have 8 neighbors, of which state will affect the state of the surrounded cell. For the given initial states of all the cells of the planar grid, the following rules should be used:

1. If the surrounded cell has 3 neighbors of which grid is black, then, no matter what its initial state was like, it has to be changed to black;

2. If the cell happens to have 2 neighbors whose state is black, it will maintain its initial state unchanged.

3. Otherwise, no matter what its initial state was like, it has to be changed to white.

If the planar grid is seen as a 1-bit image, then Conway's game will be a kind of special image matrix transformation. Figure 4.12 shows a diagram of the multiplied process of Conway's game [57].

(a) Initial state. (b) Iteration by one step. (c) Iteration by two steps. (d) Iteration by three steps.
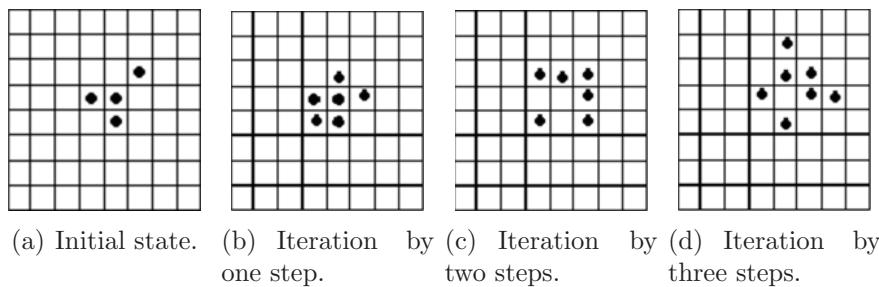
Figure 4.12: Diagram of the multiplied course of Conway's game.

Suppose that set Z represents the alive state of all the cells on the planar grid, i.e., the set of all the grids on the planar grid, set S_0 the initial state of these alive cells, i.e., the set of the grids in the initial state that become alive, and set S_i represents the set of the cells that become alive after i-th iteration, where $i > 0$. Let set $S^{(i)} = \cup_{0 \leq j \leq i} S_j$ represent all the cells, which had become alive after the i-th iteration. The algorithm can be defined as follows:

1. Establish the correspondence between the cells and pixels of a given image;

2. For the initial state S_0 of a planar grid, arrange, in order of scan lines, the pixels corresponding to the alive cells into the space of the coordinates of scrambled image;

3. When reaching the i-th steps of iteration, arrange, in order of scan lines, the pixels corresponding to the alive cells in the set $S^{(i)} - S^{(i-1)}$ into the space of the coordinates of scrambled image;

4. If the iteration stops at the n-th step, arrange, in order of scan lines, the pixels corresponding to the cells in the set $Z - S^{(N)}$ into the space of the coordinates of scrambled image.

The initial state S_0 of the planar grid can be generated by pseudo random numbers. Figure 4.13 shows different scrambled results corresponding with different initial states.
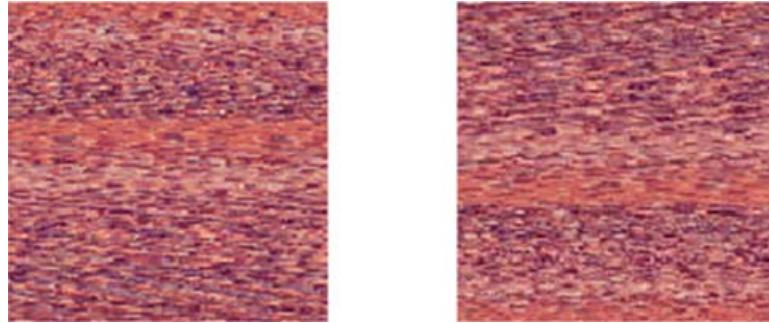
Figure 4.13: Digital image scrambling based on Conway's game.

4.3   Audio Scrambling

Since digital audio is described by signal amplitude, frequency, phase [58] and time. Therefore, these parameters are employed for audio scrambling. The basic principle of these approaches is that the bandwidth of the digital signal is not changed during the audio scrambling process [59].

Audio scrambling aims to minimize the residual intelligibility of the original audio and control the access to authorized users only. It is similar to but not a direct application of normal cryptographic techniques. The main advantage is that scrambled audio files are still legal audio files which can be played by the corresponding players.

4.3.1 Audio Scrambling in the Temporal Domain

Audio scrambling based in the temporal domain usually segment audio signals into many pieces, these pieces have been scrambled by using different parameters based on the temporal window. In each window, audio signals are scrambled in each frame. This window based scrambling improved the redundancy of scrambled audio. This approach can reach trade-off between secure degree and quality.

Audio scrambling based in the temporal domain usually adopts the uniform scrambling transformation:

$T_{k_1 m_1} = [t_{ij}]_{(N-1) \times (N-1)}, i, j = 0, 1, 2, \ldots, N-1$, k_1 and N are prime numbers:

$$t_{ij} = \begin{cases} 1 & j = (k_1 i + m_1) \bmod N \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

Its inverse transformation is:

$$T_{k_2 m_2} = T_{k_1 m_1}^1, (k_1 m_2 + m_1) \bmod N = 0, (k_1 k_2) \bmod N = 1 \tag{4.15}$$

In practice, in order to enhance the key space, the following transformation is taken:

$$T = \sum_{m_1=0}^{N-1} a_{m_1} T_{k_1 m_1} \tag{4.16}$$

Its inverse transformation is:

$$T^{-1} = \sum_{m_2=0}^{N-1} a_{m_2} T_{k_1 m_2} \tag{4.17}$$

4.3.2 Scrambling in the Frequency Domain

Scrambling in the frequency domain is committed to the frequency. In audio scrambling, in order to prevent an increase in size of the audio spectrum, it normally does not scramble the high frequency part of the signal. Therefore the following transformation is taken:

$$M = \begin{bmatrix} X & X & 0 & X \\ X & X & 0 & X \\ 0 & 0 & I & 0 \\ X & X & 0 & X \end{bmatrix} \tag{4.18}$$

where I is $N_2 \times N_2$ is unify matrix, X is the standard $N_1 \times N_1$ swapping matrix, 0 is all zero $N_2 \times N_2$ matrix, $N = 2N_1 + N_2 + 1$. The matrix is symmetric, it is useful in encoding and decoding.

4.3.3 Joint Scrambling

Joint scrambling combines temporal scrambling and frequency scrambling together. This is more powerful than only using one domain in scrambling. The following segment introduces audio scrambling based on integer rings:

In the integer remain ring z_m modeled with M, $x(q) \in Z_m (q = 0, 1, \ldots, N - 1)$ the transformation operates on the sequence $x(0), x(1), x(2), \ldots, x(N - 1)$ is:

$$x(k) = \sum_{q=0}^{N-1} x(q) a^{qk}; k = 0, 1, \ldots, N - 1; a \in Z_m \qquad (4.19)$$

It inverse transformation is:

$$x(q) = N^{-1} \sum_{k=0}^{N-1} x(k) a^{-qk}; q = 0, 1, \ldots, N - 1; a \in Z_m \qquad (4.20)$$

This transformation has the property of convolution.

4.3.4 Progressive MP3 Audio Scrambling

With the wide-spread use of the MP3 audio format, the need for developing a new scrambling algorithm that can work in the compressed domain has become important [60]. Past research is based on permutation in either the frequency or time domain [61] for normal audio data. However, not much work has been done in the compressed domain. MP3 is mainly used for online music distribution. There is a need for allowing the music owners to have flexible control over their music. An algorithm that allows them to provide music of varying quality to different users and which, at same time, is able to protect the copyright for the MP3 files can be very useful.

In this section, a progressive algorithm is introduced that does multiple rounds of scrambling of MP3 audio. The MP3 outputs can be reconstructed at different levels of quality based on the number of keys provided and the rounds of scrambling performed.

The idea behind this technique is to scramble the Huffman code words in the MP3 file. However, instead of directly shuffling the Huffman code word table, the actual MP3 data is operated upon, this is to eliminate any repetitious values, i.e. one value in the original audio will be shuffled to different output values, which will break the statistical correlation between the source audio and the output. This is very important for thwarting statistical cryptanalysis.

Progressive quality levels are attained by conducting multiple scrambling. A few rounds of scrambling are required and the quality degrades gradually with each scrambling. By performing the inverse descrambling processes, audio with different qualities will be generated.
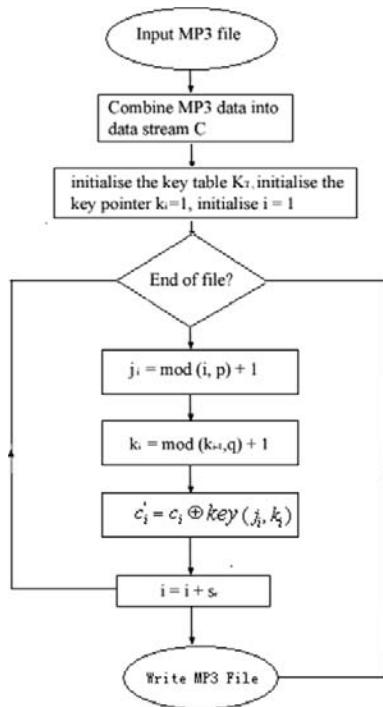
Figure 4.14: Flowchart of proposed algorithm for MP3 audio scramling.

The procedure of MP3 audio scrambling/descrambling is explained in Figure 4.14. Given an MP3 audio, the header frame is ignored and only the audio data is operated upon, $C = c_i, i = 1, 2, \ldots, n$, n is the length of the audio stream. A sample rate s_γ needs to be predefined before each round of scrambling. To make the quality degradation linear, the sampling rate $s_\gamma(l)$ needs to be set at an exponential increment rate for different rounds.

$$s_\gamma(l) = a^l, a \in Z \tag{4.21}$$

where l is the level number, $l = 1, 2, \ldots, L$. L is the total number of rounds of scrambling to conduct. Also a K_T will be initialized; the format of K_T is shown as follows:

$$K_T = (key(i,j))_{p \times q} \tag{4.22}$$

The key table generation can make use of special techniques to ensure the uniqueness and randomness of the key tables generated. For the keys used in this scheme, the Arnold matrix [62] is employed to generate random $p \times q$ matrix using an input random generated seed, here $p = q$. The Arnold matrix is shown as follows:

$$A = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & 2 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 2 & \ldots & p \end{bmatrix} \tag{4.23}$$

First, generate a random matrix $R = (r_{i,j})_{p \times q}$

$$r_{i,j} = rand(range), i = 1, 2, \ldots, p; j = 1, 2, \ldots, q \tag{4.24}$$

where $rand(\cdot)$ is a function to generate random number sequence based on a seed $range$. Then generate K_T' by doing a matrix product of the Arnold matrix [62]:

$$K_T = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & 2 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 2 & \ldots & p \end{bmatrix} \cdot R (\mathrm{mod}\ p) \tag{4.25}$$

The Arnold matrix is used to transform the keys in the table to the different kinds of choices for different users. The transformation based on this matrix can ensure the validity of the keys since they are bounded.

For the i-th data item c_i in the original audio, a modulo division is performed of the current position i with the row length p of the key table and the remainder decides the row number j of the key to be used.

$$j_i = \mathrm{mod}(i, p) + 1 \tag{4.26}$$

The column number k_i will cyclically range from the first column to the maximum column number q.

$$k_i = \mathrm{mod}(k_{i-1}, q) + 1 \tag{4.27}$$

After the key $k_i = key(j_i, k_i)$ is selected, an XOR is performed with the MP3 audio datum c_i. XOR is chosen due to the speed and simplicity of implementation, since it is its own inverse. The changed datum c_i' is written into the corresponding position of the output stream:

$$c_i' = c_i \oplus key(j_i, k_i) \tag{4.28}$$

where $key(j_i, k_i)$ is the element in matrix K_T, $\oplus$ is the bit-wise XOR. This procedure is repeated until the end of the data stream and the output data stream is also an MP3 stream. The keys used will be stored and they will be distributed to the authorized consumers for descrambling.

The descrambling process is exactly similar to the scrambling procedure. For descrambling level l, a MP3 file generated from descrambling level $l - 1$ will be used as an input. The same process is applied and using the same K_T that has been used in the scramble round l.

With the implementation of the scrambling process as described above, the quality of the output MP3 file depends on how many levels of descrambling involved. For example, during scrambling if an MP3 audio has been scrambled 5 times using 5 different keys, a user needs to have all the 5 keys and perform 5 rounds of descrambling before they can perfectly restore the original audio. If the user has only 4 keys or performs only 4 rounds of descrambling, he will get an audio of approximately 80% of the original quality.

Figure 4.15 depicts the waveform for an audio file at different scrambled levels. Figure 4.15(a) is the waveform of source MP3 audio, the audio is scrambled five times and Figure 4.15(f) is obtained. Figure 4.15(f) is descrambled to get Figure

4.15(e), which is of slightly better quality, further descrambling produces Figure 4.15(d) to Figure 4.15(b) by performing yet more additional rounds of descrambling. The waveforms show that the output audio clips are gradually reconstructed and eventually an audio signal will be obtained that is exactly same as the source.

(a) Original MP3 audio. (b) Scrambled audio at level 1. (c) Scrambled audio at level 2.

(d) Scrambled audio at level 3. (e) Scrambled audio at level 4. (f) Scrambled audio at level 5.
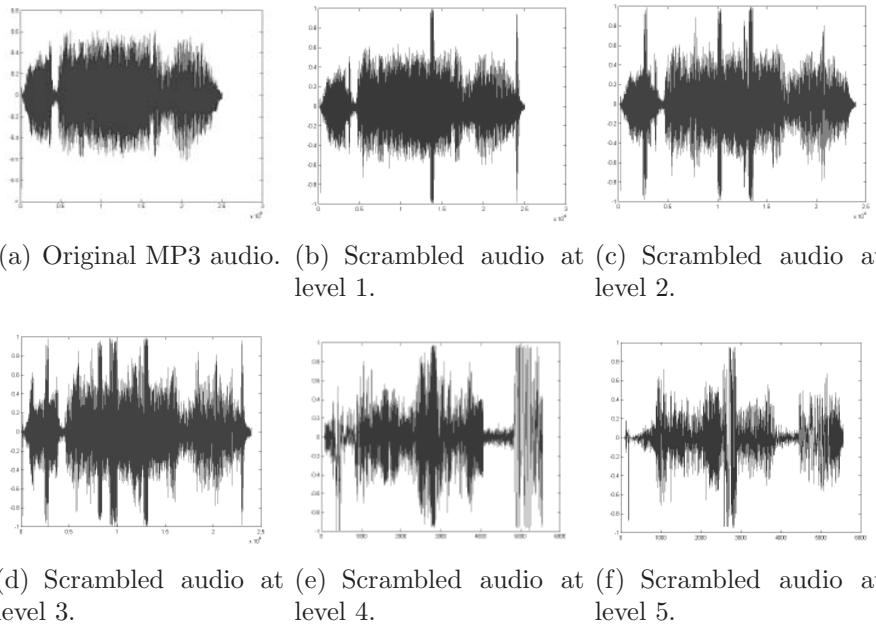
Figure 4.15: Original and scrambled MP3 audio waveforms at different scrambled levels.

4.4   Summary

In this chapter, different aspects of media scrambling were discussed, from image scrambling to audio scrambling. The principles of multimedia scrambling were also introduced along with how to analyze the scrambled media data. Generally all semantic multimedia data has the same security problem, if the content needs to be protected, multimedia scrambling will play a pivotal role in doing so.

Chapter 5

Digital Surveillance

Due to the decreasing costs and sizes of video cameras, the use of digital video based surveillance as a tool for real-time monitoring is rapidly increasing. In this chapter, the situation of multiple surveillance cameras is taken into account and the experiential sampling technique is utilized to decide which surveillance video stream should be displayed on the main monitor. This can help tremendously in reducing the amount of concentration required for multiple monitor situations.

5.1   Overview

Surveillance is being increasingly used for traditional and non-traditional security applications such as monitoring of shopping malls and ATMs as well as industrial supervisory use and city road networks [63]. The decreasing costs coupled with rapid miniaturization of the video camera have enabled its widespread use on highways, airports, railway stations and on-board vehicles. The recent trend of coupling video cameras to cell-phones will only accelerate this trend. Therefore, research in video surveillance is moving into the mainstream with the focus on day-to-day applications and uncontrolled outdoor scenarios. It is moving away from mere data collection with manual observation to intelligent analysis of events and actions at a semantic level without the intervention of humans.

5.2   Adaptive Video Monitoring

There is a growing interest in the use of video surveillance techniques aided by the decreasing cost of sensors and an increasing set of vulnerabilities. Video surveillance can potentially provide a cost-effective means of eliminating or mitigating potential

security breaches. Video surveillance has applications in offices, stores, public spaces and homes [64, 65]. This chapter will provide an incite into experiential sampling techniques in a feedback control system for automatic adjustment of the parameters of a fixed video camera. The panning and zooming parameters of a video camera will be controlled during this setup.

Motivation for studying the automatic adjustment of camera parameters stems from the fact that a camera for video surveillance is usually not manipulated after its installation. Regular calibration of such cameras for object tracking is desirable but not feasible if done manually. As a result surveillance videos often are of low quality due to changing ambient conditions especially for outdoor settings. High quality videos in focus with appropriate positions for monitoring moving objects require the fixed surveillance cameras to capture and track those moving objects at its optimal framing position with correct zoom and pan settings. Thus the automatic setting and control of such video cameras is a difficult but extremely useful task. Manual control of cameras via internet, wireless [66] or remote radio channels is a feasible but tedious operation. The aim is to have precise control of surveillance cameras via these channels automatically in order to obtain a superior quality surveillance video.
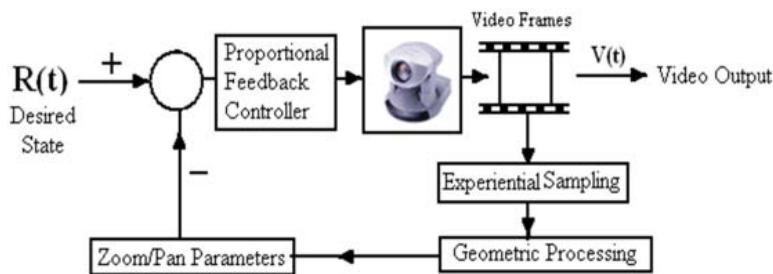
Figure 5.1: Feedback Control System for Surveillance.

Figure 5.1 is the proposed framework to be discussed throughout this chapter. Several techniques for the detection and tracking of moving objects have been investigated in the literature. The experiential sampling technique is used because it effectively takes advantage of the context information of surveillance videos. For example, the future motion of moving objects can be predicted by using past knowledge that has been learned by the system. Thus, the computation is performed in an experiential environment which is captured by both the sensor and attention samples. The number of attention samples used is therefore a function of the environment as well as past experiences. Consequently, experiential sampling creates a dynamical solution for real-time object detection and tracking.

In order to obtain robust and accurate results, video surveillance system should have adaptability in order to respond to the changes in the monitored environment. There

can be potentially many variations in the monitored environment. These variations mainly come from two aspects: geometric aspects (location, size) of the observed objects and the visibility quality (lighting, contrast) of the monitored environment.

Currently available video cameras can be manipulated to track the observed object of interest by performing panning and zooming operations. It is also possible to adjust to the visibility of the environment by automatically controlling camera parameters such as brightness and contrast. Therefore, it would be extremely useful to systematically manipulate these camera parameters so as to adapt to the changing surveillance environment.

The adaptive monitoring method considers both geometric aspects (automatic panning and zooming) in order to follow the movements (position, size) of the monitored object and visibility aspect. Since the adjustment of visibility parameters is similar to previous work about automatic adjustment video quality for home video in [67]. The automatic adjustment of the visibility parameters are not experimented on, as opposed to only including it within the framework.

As previously mentioned, in a surveillance system, there are two types of variations in the environment: geometric aspects of objects and the visibility. In order to adaptively monitor the environment, the basic experiential sampling algorithm should be embedded into a feedback control system. The "state" variables of this feedback control system are precisely the two factors of the environment which impact the surveillance task. Thus, the state vector of the feedback control system is represented by:

$$X(t) = \{Geom(t), Vis(t)\} \tag{5.1}$$

where $Geom(t)$ captures the geometric aspects of the surveillance object of interest and $Vis(t)$ represents the visibility aspects of the object in the ambient environment. The state variables of this system (geometric aspects of object and the visibility of the environment) have to be estimated based on the previous time-instant attention samples and sensor samples. After obtaining those two factors, the camera parameters can be adjusted by providing the appropriate feedback According to:

$$Geom(t) = \{zoom(t), pan(t)\} \leftarrow \{AS(t-1)\} \tag{5.2}$$

$$Vis(t) = \{brig(t), contr(t)\} \leftarrow \{SS(t-1)\} \tag{5.3}$$

where $Geom(t)$, the geometric parameters at time slice t, this includes the zoom parameter $zoom(t)$ and pan parameter $pan(t)$. $Vis(t)$, the visibility parameters to be tuned includes $brig(t)$ the brightness and $contr(t)$ contrast parameters.

Having characterized the surveillance environment by these two state variables, the overall feedback control algorithm with experiential sampling embedded can be described as follows:

Algorithm 1: Feedback control algorithm with experimental sampling.

    Input: Surveillance video stream from a camera;
    Output: Surveillance object, camera feedback;

1   **begin**
2     $t = 0$;
3     $X(t) \leftarrow \{\text{Geom}(t), \text{Vis}(t)\}$;
4     $\{SS(t)\} \leftarrow$ uniform sampling;
5     $\text{Asat}(t) \leftarrow$ sum of $\{SS(t)\}$;
6     $\text{Vis}(t+1) \leftarrow \{SS(t)\}$ [visibility feedback];
7     $N_s(t) \leftarrow \text{Asat}(t)$;
8   **if** $N_s > 0$ **then**
9 | go to step 7;
10   **end**
11    $t = t + 1$; goto step 3;
12    $\{AS(t)\} \leftarrow$ important sampling from $\{SS(t)\}$;
13   **foreach** AS **do**
14 | perform analysis;
15   **end**
16    $\text{Geom}(t+1) \leftarrow \{AS(t)\}$ [geometric feedback];
17    $t = t + 1$; goto step 3;
18 **end**

After the geometric parameters of the object at time t are obtained, the panning and zooming parameters of the camera can be determined for the next time instant $t + 1$, and thus can use them to control the video display. This operation is the geometric processing operation. The procedure is quite similar to the adjustment of the visibility parameters. The geometric aspect is described in detail now.

Geometric processing aims to obtain the optimal framing for the surveillance object of interest. Attempts are made to frame a moving object (e.g. human face in this implementation) based on the attention samples. The optimal zooming factor and translation vector are obtained to provide the feedback to the surveillance camera. The proportional feedback control strategy [68] is used for doing the corrective feedback. At first, the desired state of the surveillance system is fixed. If the objects of interest are not framed properly (which will be indicated by the attention samples), the target state will be reached by appropriate modification of the pan and zoom parameters of the camera. For instance, if the tracked object is at the bottom of the video frame, the pan parameter of the camera is controlled and the frame is moved upwards. Similarly, if the tracked object is at the top of the video frame, the frame is panned downwards. Simultaneously, if the objects are not at the desired zoom level, the zoom parameter will be controlled to obtain the video with the appropriate size of the objects. The system settles down to a steady state after a period of transition when the feedback control system makes the appropriate corrective actions.

For this system, the desired state of the system is fixed as follows: (a) the centroid of the object of interest should be located at the center of the video frame and (b) The ratio of the sides of the bounding box of the object of interest to the sides of the video frame should be 0.618. From a human visual system point of view, the object of interest at this framing state will exhibit the best observability. Such a video frame can then be suitably displayed on a monitor to be seen by a human operator. Based on the attention samples in the previous time instance, the two parameters are computed as follows:

$$O_c = \{AS(t-1)\}_c \tag{5.4}$$

$$AR = Area_i / Area_{\{AS(t-1)\}} \tag{5.5}$$

where the object centroid O_c is approximated by the centroid of all attention samples (denoted by $\{AS(t)\}_c$) while AR represents the area ratio of the total image area ($Area_i$) to the area of attention samples ($Area_{\{AS(t)\}}$).

The definition of $pan(t)$ is the camera pan position at time t. Therefore the new $pan(t)$ in step 16 of the feedback control algorithm can be formalized as:

$$pan(t) = pan(t-1) + \alpha(O_c - pan(t-1)) \tag{5.6}$$

where α is the factor to control the panning speed. $\alpha = 0.3$ was used in the experiments. The definition of $zoom(t)$ is the object size at time t.

Notice that the error between the desired state of the system and the actual observed state of the system is used as the corrective stimulus to steer the system towards the desired state. This is basically due to the proportional feedback control strategy used.

5.3 Video surveillance using multiple cameras

Without appropriate management, the huge volume of filed video data will be impossible to analyze for events of interest. Video surveillance chiefly deals with spatio-temporal data which have the following attributes:

- They possess tremendous volumes;

- The data is dynamic with temporal variations with a resultant history;

- Some data can be live with real-time processing and filtering requirements;

- It does not exist in isolation - it exists in its ambient context with other data.

If a video surveillance technique does not fully consider the above attributes, it can lead to tremendous computational inefficiency as well inflexibility in the processing. The inefficiency stems from the inability to filter out the relevant aspects of the data and thus considerable resources are expended on superfluous computations on redundant data. If the ambient context is ignored, the technique cannot react to the changing environment. Thus, the surveillance processing cannot adapt itself to the task at hand.

On the other hand, there is solid evidence that humans are superb at dealing with large volumes of disparate data using their own senses. The human visual system is particularly adept at understanding the surrounding environment at appropriate accuracy quite efficiently. This is due to many factors: the excellence of the physical visual system, the richness of fusion information from perception, implicit understanding of every visual object, and the common understanding of how the world works. These attributes in the experiential environments [69, 70] play an important role for the human visual perception in understanding the visual scene accurately and quickly. Some of these strategies would be incorporated into the techniques for

video surveillance. The main idea in this chapter is to enable surveillance techniques with an ability to "focus" precisely on the data of interest while being simultaneously sensitive to the current context as well as the assimilated past experiences.

In order to achieve this, a novel technique called experiential sampling is utilized, i.e., sampling the data in the experiential environment. The main theoretical technique has been introduced initially by Wang and Kankanhalli [71]. This experiential sampling technique is applied to the problem of video surveillance and demonstrates its usefulness for this problem. The basic idea is to sense the contextual information in the experiential environment in order to build a sampling based dynamic attention model to maintain the focus towards the interest of the current surveillance task. Only the relevant samples are considered for the surveillance analysis. These samples succinctly capture the most important data. Moreover, the past samples influence future sampling via feedback. This mechanism ensures that the analysis task benefits from past experience.

The experiential sampling technique can be utilized for many aspects of video surveillance such as object (face or vehicle) detection, object recognition and object tracking. For example, intruders can enter only from the boundary of the scene. Therefore, when there are no intruders in the scene, the analysis task for surveillance should focus on the boundary. If there is an intruder, the focus of attention should evolve to follow the person. These experiences can be easily modeled by using the experiential sampling technique. The results of multiple cameras video surveillance are presented.
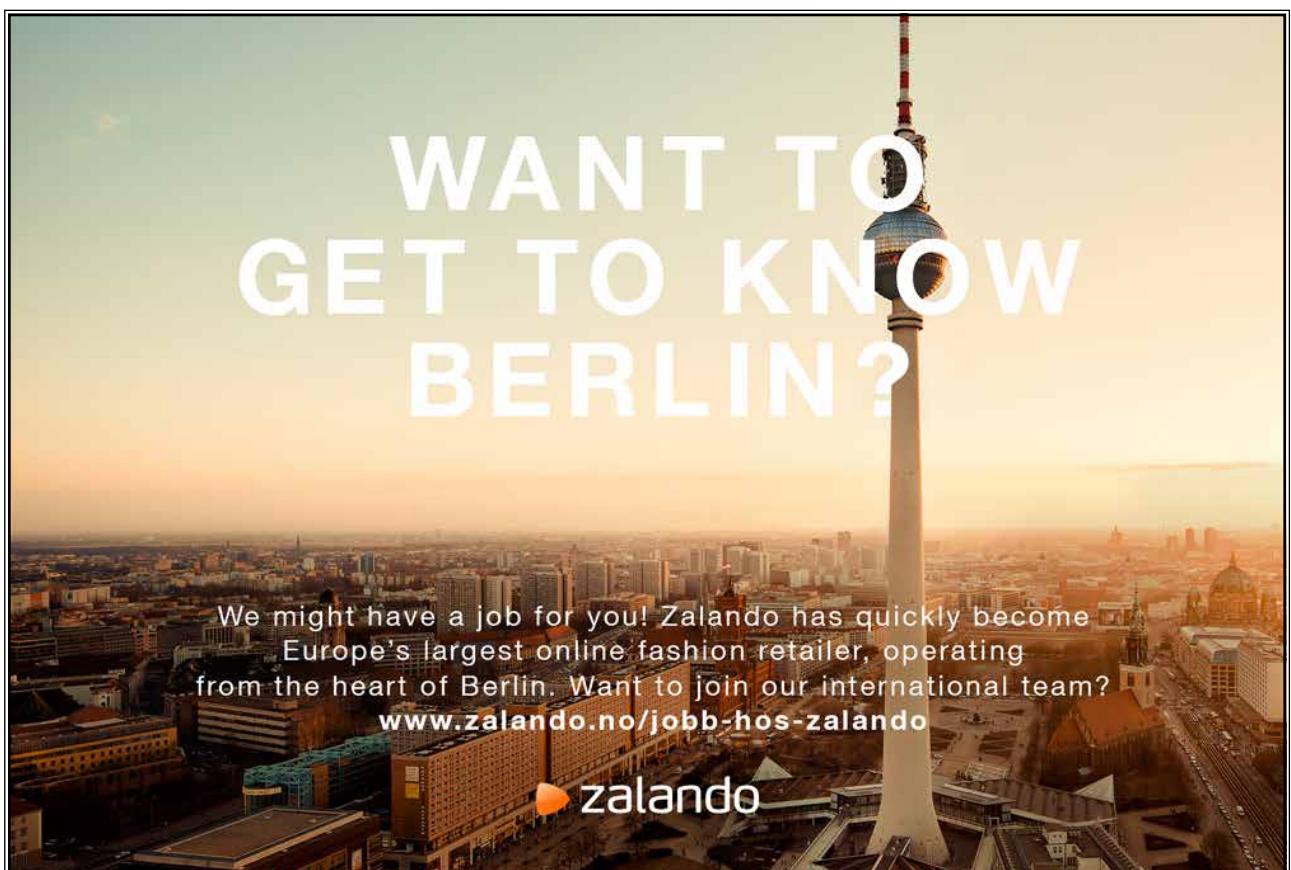
As a general analysis framework for video surveillance, this proposed approach can be used for a variety of video surveillance tasks, especially real-time applications. As a test case, a multiple camera surveillance scenario has been developed. This scenario is described first before presenting the experimental results.

In the real-world usage of video surveillance, multiple cameras are utilized in a wide spectrum of applications for the purpose of monitoring. As a result, multiple monitors are used to display the various output video streams. Manual monitoring of multiple screens is indeed tedious and operators are prone to fatigue which can have disastrous consequences. Thus, it would be extremely useful to reduce the data from multiple screens into one main screen for monitoring. This is reasonable since usually only one screen is of active interest which engages the attention of the operator. Even if there is no manual observer, finding the most relevant camera data stream is useful for automated analysis such as object detection. Thus in this section, an algorithm is developed to analyze the importance of the video frames from the multiple video cameras by using the experiential sampling technique.
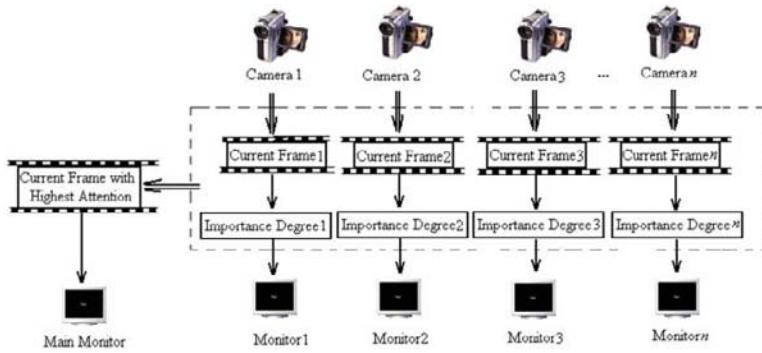
Figure 5.2: The setup for multiple camera surveillance.

In this multiple camera surveillance scenario, there are p number of cameras C_k. The experiential environment can be redefined as:

$$e_t = \{C_1(t), C_2(t), \ldots, C_p(t)\} \text{ and } C_k(t) = \{S_k(t), A_k(t)\} \tag{5.7}$$

where $A_k(t)$ and $S_k(t)$ are the attention samples and sensor samples of the camera $C_k(t)$ at time slice t. By this definition, the algorithm introduced in the previous section is used to acquire the attention saturation of the motion activity. It is assumed that the motion attention saturation precisely reflects the importance of current video frame. At any point in time, the output of one camera is designated as the relevant camera whose output is displayed on the main monitor. The actual camera picked dynamically changes based on the data. The strategy is illustrated in Figure 5.2. In Figure 5.2, several cameras are fixed and their current frames are

extracted for analysis. The attention saturation of these frames are computed and compared. Only the frame with the maximum attention saturation will be extracted and displayed on the main monitor.

The algorithm for multiple camera video surveillance can be described as follows:

1. Collect all the video frames at this time instance;

2. Compute their attention saturation;

3. Find the most important frame;

4. Display the most important frame on the main monitor;

5. Continue till the end.

The advantage of this proposal is that the most important frame is extracted at any given instance and rendered on the main monitor which greatly reduces the manual monitoring efforts. Note that this can be easily generalized for the n cameras and m monitors situation where $m < n$.
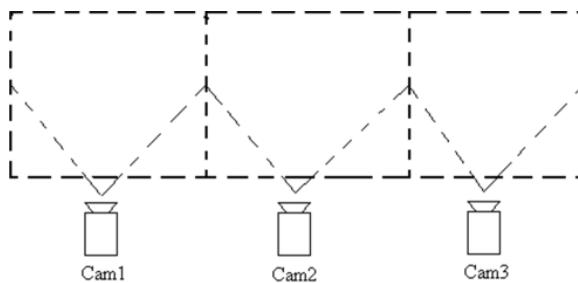
(a) Diagram of camera position.
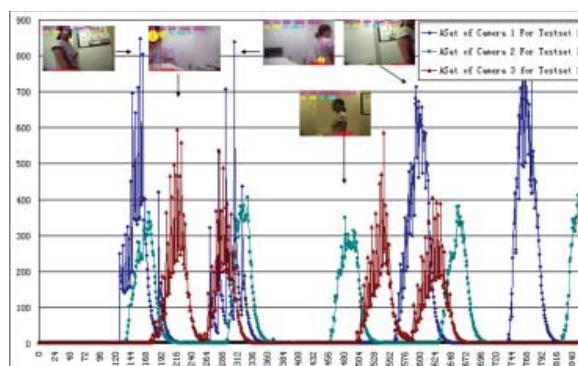
(b) Attention saturation.

Figure 5.3: The attention saturation of surveillance videos (Group 1).
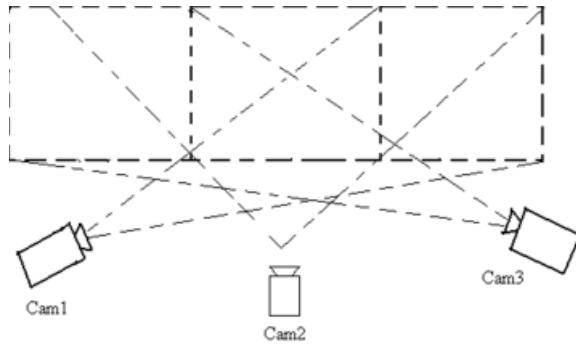
(a) Diagram of camera position.
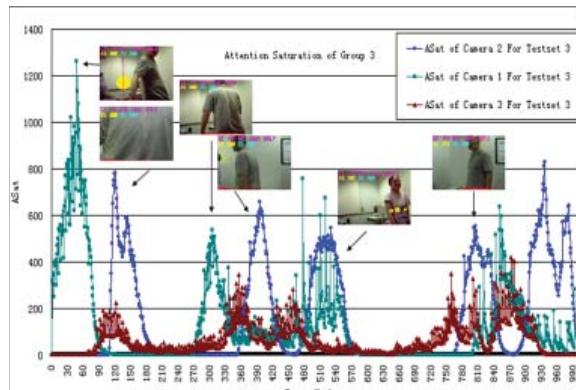
(b) Attention saturation.

Figure 5.4: The attention saturation of surveillance videos (Group 2).

Figure 5.3 and Figure 5.4 represent two types of experiment. Individual attention saturation of each frame in these three videos is calculated. The important frames are extracted, namely the ones with the highest attention saturation to compose a new video, this new video will be displayed on the main monitor for the operators. Figure 5.3(a) is a diagram for the positions of three video cameras. Each camera covers only a limited field of view, and these viewing regions have no overlap. Figure 5.3(b) is the resultant diagram for the attention saturations of the three video streams. It can be clearly seen that the peaks occur at different times in the different cameras. A peak corresponds to the fact that someone is detected to be walking past that camera. Thus the main monitor can display the output of the camera which is active at any instance of time (instead of constantly watching three monitors).

Figure 5.4 shows another group of results. The three cameras have some overlaps in the field of view as shown in Figure 5.4(a). In this case, the person walking can be seen on many monitors simultaneously. In this case, it becomes very difficult for a manual operator to decide which monitor to watch. However, by using attention

saturation for experiential sampling, the peaks can be easily determined as shown in Figure 5.4(b). This information can be utilized to appropriately display the information on the main monitor.

5.4   Multimedia Simplification for Video Monitoring

Multimedia simplification is related to but is very different from summarization. It attempts to pick the semantically most salient aspects of the media streams. Simplification focuses on extracting the semantic saliency of the multimedia data while summarization attempts maximal data reduction with minimal loss in semantics. Thus, simplification is always semantically lossy which preserves only the most salient aspects for a given data size constraint. It helps obtain the most compact representation to convey the semantic saliency. On the other hand, summarization tries to minimize this semantic loss while maximizing the data compactness. In other words, the summary tries to provide a gist of the entire content - the salient aspects as well as some idea of the non-salient aspects. It must be noted that a drastic amount of summarization will asymptotically reduce to simplification since the salient aspects will be preserved at the cost of the non-salient aspects.

However, there is another crucial difference between simplification and summarization in the case of multimedia data. For multiple correlated streams, summarization faithfully tries to preserve the individual stream semantics without trading off one against the other. On the other hand, simplification on correlated streams can potentially do cross-modal tradeoffs. For instance, the video stream could be totally discarded if the audio stream can completely convey the salient aspect (e.g. the current score of a live game).

Pictures are two-dimensional signals. Experiential sampling was first used to reduce the redundancy in surveillance video [72]. The experiential sampling technique is employed on pictures to obtain the most salient region of the picture. With the availability of high resolution cameras, images tend to be detailed and huge. It may be possible to reduce the resolution via sub-sampling, which can however result in the loss of details, which may not be desirable. Hence, it would be better to have a cropped but detailed rendition of the most salient region of the image.

Simplification of Still Pictures

First, the gradients of the image is calculated using [73]:

$$\partial I = \left(\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y} \right) = (I_x, I_y) \tag{5.8}$$

where $I(x,y), x = 1, 2, \ldots, W; y = 1, 2, \ldots, H$, W and H are width and height of the image respectively. The intensity changes are computed to discard the low-gradient regions.

$$\partial I = \left(\frac{I(x + \Delta x, y) - I(x,y)}{\Delta x}, \frac{I(x, y + \Delta y) - I(x,y)}{\Delta y} \right) \tag{5.9}$$

Δx and Δy are the step lengths in the X and Y direction respectively. The sensor samples of the ES algorithm are $SS(t) = \{\partial I(x,y)\}$ where W_t is a candidate window. If $|\partial I(x,y)| > \tau$, then $I(x,y)$ belongs to the salient region. The attention saturation is $A_{sat}(x,y) = \{I(x,y), |\partial I(x,y)| > \tau_1, \tau_1 > 0\}$. The attention samples are determined by the ratio between the salient regions and the non-salient regions, namely:

$$AS(t) = \{\partial I(x,y), \frac{N_{os}(t)}{N_{oN}(t)} > \tau_2, (x,y) \in W_t\} \tag{5.10}$$

where $N_{os}(t)$ and $N_{oN}(t)$ are the numbers of salient and the non-salient regions. Only the important region $\Omega = \{W_t : \frac{N_{os}}{N_{oN}} > \tau_2, \tau_2 > 0\}$ will be selected and presented on the mobile device, W_t is t-th window on the picture. Basically, this technique is based on image segmentation which utilizes image block fusion and multi-level selection based on thresholds.

Simplification of Motion Pictures

For the motion pictures, the salient portions need to be extracted from the important frames. It is assumed that the entire set of full pictures for an event are obtained [74]. The most important frames are selected from the entire sequence. For a single picture, the earlier technique can be used to perform still picture simplification. However, this will lead to a problem. For a sequence of pictures in a video, different frames may be cropped in different places with different aspect ratios. So if a slide-show of such a sequence is run, from individually simplified frames, it is likely that some jitter can be observed. Hence, for motion pictures, the entire sequence of frames must be considered in order to eliminate jitter. For this purpose, the motion trajectory is computed of the single-frame simplification results with respect to the original frames. Then a trajectory smoothing operation is performed based on either the mean trajectory technique or the Bezier curve fitting technique as in [67]. Given a motion picture set $M_p = \{P_1, P_2, \ldots, P_n\}$, the gradient of the motion pictures is given as follow:

$$\partial P_t = \left(\frac{\partial P_t(x,y)}{\partial t}, \frac{\partial P_t(x,y)}{\partial x}, \frac{\partial P_t(x,y)}{\partial y} \right) \tag{5.11}$$

For discrete frames, it reduces to:

$$\partial P_t = \left(\frac{\partial P_{t+\Delta t}(x,y)}{\Delta t}, \frac{\partial P_t(x+\Delta x, y)}{\Delta x}, \frac{\partial P_t(x, y+\Delta y)}{\Delta y} \right) \tag{5.12}$$

where $t = 1, 2, 3, \ldots, n$; $\Delta t, \Delta x$ and Δy are the step length in different directions. For the salient regions of each picture $P_i = \{O_{i1}, O_{i2}, \ldots, O_{im}\}$ selected from Eq. 5.10, the importance of each object is $|O_{ij}|$. The importance of that frame among motion pictures is measured by $P_i = \max |O_{ij}|$. Using this information, salient frames can be picked as well as the salient regions of every picked frame. The motion trajectory can then be computed which is then smoothed. The final set of salient regions is then obtained.

Audio simplification aims to find the most salient content in an audio clip. The saliency is determined by the energy and spectrum of the audio content. It is based on intra- and interframe analysis. The experiential sampling technique is used to discard the low attended audio frames. Consider an audio clip: $\gamma = \{v_i, i = 1, 2, \ldots, n\}$ where v_i are audio frames; for each frame, the samples are $v_i = \{\Psi_{i,1,1}, \Psi_{i,1,1}, \ldots, \Psi_{i,1,1}, \Psi_{i,j,m}\}$, i is number of this frame, j is the number of channels and m is number of samples per frame. The size of each frame in bytes is given by:

$$L = \frac{i \times j \times w}{8} \tag{5.13}$$

where w is the number of bits allocated for each sample. The sensor samples $SS(t)$ or audio experiential sampling (using the ES algorithm) are calculated as:

$$SS(i) = \left\{ v_i, \sum_{i=1}^{n} |\Psi_{i,j_0,k} - \Psi_{i,j_0,k-1}| > \tau_1 \right\} \tag{5.14}$$

$$A_{sat}(i) = \left\{ v_i^1 \in SS(i), \sum_{i=1}^{n} |\Psi_{i,j_0,k} - \Psi_{i,j_0,k-1}| > \tau_2 \right\} \tag{5.15}$$

The attention samples can now be computed:

$$AS(i) = \left\{ v_i^2 \in A_{sat}(i) : |\max_k(\Psi_{i,j_0,k}) - \min_k(\Psi_{i,j_0,k})| > \tau_3 \right\} \tag{5.16}$$

Where $\tau_i > 0, i = 1, 2, 3$ are the thresholds for determining the different degrees of salience.

In the scenario of family care, the sensors installed at home will transfer the visual and audio signals to a home server. Only the relevant simplified information is integrated and transferred to the home server. Once an alarm is triggered, the home server sends the relevant data to the MMS server. The server processes the multimedia data and sends the information to the mobile device. On the mobile device, the important information will be shown. If the mobile users would like to know in detail about the alarm, he can request the MMS server to send more information.

Here a motion picture sequence is shown for infant care. For the motion pictures corresponding to Figure 5.5(a), two pictures are obtained having high saliency among the sequence. The obtained pictures in Figure 5.5(b) are useful in monitoring the baby's actions especially to check for face laceration with their fingernails. The two important pictures appear at frame numbers 198 and 286 from the total of 366 pictures. Figure 5.5(c), Figure 5.5(d), Figure 5.5(e) and Figure 5.5(f) are the results after simplification at important degrees of 70%, 65%, 60%, and 55% respectively.

5.5   Summary

In this chapter, media surveillance has been discussed. Three problems were looked at: adaptive camera automatically control based on content feedback, multiple camera surveillance environment, and multimedia simplification based video monitoring.

The theoretic base behind these applications is experiential sampling, which can predict the development of the scene according to the events happened. This is a very important part in experiential computing and provides a very useful and effective way of using media surveillance.

(a) Motion pictures for the new born baby.

(b) Two important pictures among the picture sequence ($< 95\%$).

(c) Three important pictures among the picture sequence ($< 70\%$).

(d) Four important pictures among the picture sequence ($< 65\%$).

(e) Five important pictures among the picture sequence ($< 60\%$).

(f) Seven important pictures among the picture sequence at 8, 66, 130, 269, 276, 198, 286 frames respectively ($> 55\%$).

Figure 5.5: The simplified motion pictures for baby care.

Chapter 6

Digital Multimedia Authentication and Forensics

Multimedia authentication and forensics are most important recent research topics. With the exponential increase of digital cameras, authentication and forensics problems become more and more prominent. In this chapter, video authentication is discussed initially, the problem of print authentication and context authentication are then taken into account, finally some technologies used in photographic forensics are presented.

6.1 Video Authentication

Video authentication techniques have experienced a tremendous rise in interest over the past few years. By definition, video authentication is a process that is used to ascertain the trustworthiness of a digital video. In other words, a video authentication system ensures the integrity of digital video, and verifies that the video taken into use has not been tampered with. The need for authenticating a digital video is shown by giving the following examples:

1. A video clip can be doctored to defame a person. On the other hand, criminals get away from being punished because the video showing their crime can't be proved conclusively in the court of law.

2. In surveillance systems, it is hard to reassure that the digital video produced as evidence is the one that was actually shot by the camera.

3. A journalist cannot prove that the video played by a news channel is trustworthy.

4. A video viewer who receives video through a communication channel can not ensure that video being viewed is really the one that was transmitted. So there is a compelling need for video, wherever it is and in whatever form it is, be authenticated before use.

A typical video authentication system is shown in Figure 6.1, In the authentication process (Figure 6.1 (a)), for a given video, the authentication algorithm processes the features extracted from the video and outputs the authentication data which is encrypted using the encryption key to form the signature. The video integrity is verified by computing the new authentication data using the same authentication algorithm and features. The new authentication data is compared with the original authentication data as shown in Figure 6.1 (b). If both match, the video is treated as authentic otherwise it is construed to be tampered with.
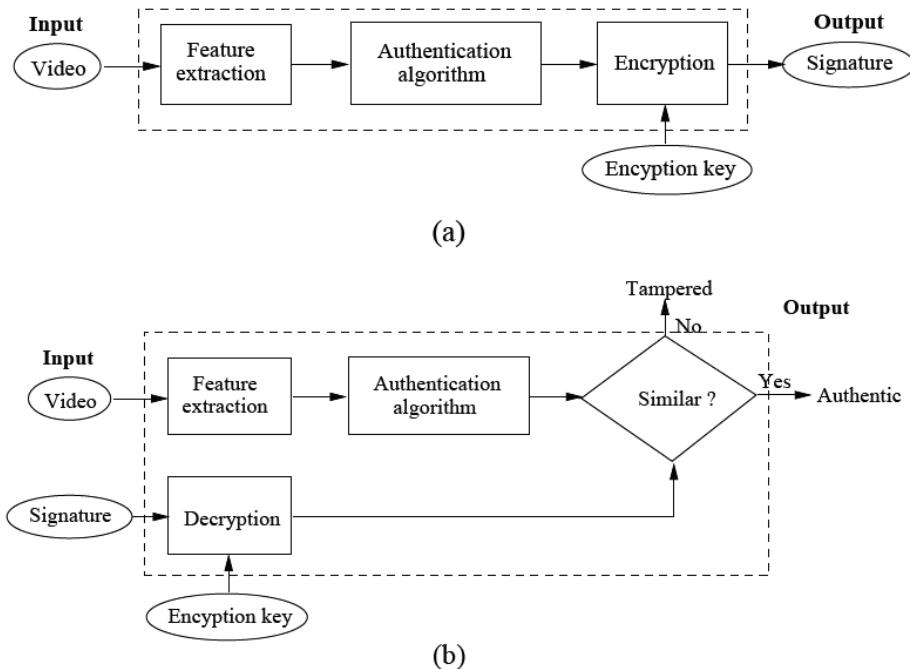
Figure 6.1: A typical video authentication system. (a) Authentication process (b) Verification process.

An ideal video authentication system, to be effective, must adhere to the following properties, such as sensitivity to alterations, localization and self-recovery of altered regions, robustness to benign operations, tolerance to some loss of information due to benign operations, compactness of authentication data, one-way property of authentication data, sensitivity against false alarm and computational feasibility.

6.2   Printed Document Authentication

Research in the authentication of printed paper documents has been growing due to its commercial potential. Although the problem has been tackled using cryptographic techniques in the digital domain [75], solutions for protecting physical documents, especially paper documents, is much less advanced [76]. Paper documents still form the basis of today's business transactions and administrative processes, and "will continue to occupy an important place in office life, but will increasingly be used in conjunction with an array of electronic tools"[77]. For that reason, authenticating printed paper documents, which is the link between electronic tools and paper documents, becomes extremely important. In the traditional paper-based world, when an authentic document is generated, it is usually signed / issued / approved by one or more authorized persons, with their signatures or seals to show the authenticity. The document with original signatures is considered to be original, authentic or legitimate. In the printed world, there are also requirements for such signatures to show the authenticity and originality of a document. Efforts towards this can be categorized into four classes:

Use of Special Material: These solutions are based on either physical means or chemical means, such as special high-resolution ($> 4000dpi$) printers not available in the open market, special papers / inks that are very sensitive to reproduce [78, 79], and hologram labels [78]. By controlling the availability of these materials, no forgery or duplication of the document is possible. However, due to the high cost of both the equipment and the efforts for controlling their use, these solutions are only used in applications which have strict security requirements, such as currency notes or bank cheques.

Fingerprints: The idea of fingerprinting is to make each copy of a document unique so that illegal copies are identifiable, or the person who made illegal copies is traceable. This idea was first introduced by Wagner in [80], and then developed for various applications. In [81], non-uniformities in disk medium are utilized as fingerprint to discourage illegal copying of files. In [82], the width of each strip cut produced by a shredder is identified as the fingerprint, which in turn is used to trace the particular shredder that has been used. As for paper documents, Metois et al. [83] have proposed an identification system based on the naturally occurring inhomogeneities of the surface of paper. A special purpose imaging device is developed to capture the texture and fiber pattern of the paper. The pattern is then registered as a unique fingerprint for later retrieval and comparison. Physical fingerprints usually 145 offer strong protection against duplication attempts. However, the medium is not content-related. Therefore, the integrity of the contents is not protected. Furthermore, the identification of typically invisible fingerprint often requires special devices. This inevitably increases the cost of the system. As a result, these methods are only used in applications which emphasize more on medium security than content integrity, such as bank cheques or printed tickets.

Digital Methods: Originating from traditional cryptography, these approaches intend to transfer digital signature onto paper documents. Such approaches include bar codes [84, 85] and information hiding (notably digital watermarking) [21, 86]. These methods add some machine readable information onto the document to serve as a digital signature. Only authorized persons have access to the private key required to generate the digital signature, so the authenticity of the document is protected. However, since the information is machine readable, it can also be copied or scanned using photocopiers or scanners. The originality of the document is not protected effectively. Digital encoding methods have been widely used in applications which require machine based authentication, such as bills, ID cards, and so on.

Visual Cryptography and Optical Watermarks: Visual cryptography utilizes secret sharing to split a graphical pattern into different pieces in a manner that the pattern becomes visible if and only if the shares are stacked together [87, 88]. By doing this, a paper document with one share printed can be validated visually using the remaining shares. Optical watermarks is an improvement over visual cryptography in terms of the ability to hide multiple layers of graphical information and enhanced visual quality with easy alignment [3]. Both visual cryptography and optical watermark have been designed for manual authentication of documents. They are most suitable in applications where the convenience of verification is important like in brand protection and printed concert tickets. However, both of these techniques cannot disprove the authenticity of a photocopy or scanned-copy of an original document. Print signature is particularly suitable for applications that re-

quire documents to be protected against unauthorized duplications while allowing the verification of the document to be convenient as well. Such applications include a shipping label, online tickets, lottery tickets and voting ballot paper.

As the laser printing technology improves, the printing resolution will become even higher. However, as long as the underlying mechanism is unchanged, it is still expected to see the random phenomenon on each copy of printed paper. This will only entail the use of microscopes of even higher resolution. This method can be readily extended to other document types such as offset-printed documents, ink-jet printed documents, or manually signed documents. It basically reduces to the task of finding unique random phenomenon in each copy of the document to be used as a signature. For example, the ink trail for each manually signed document is unique. As long as the uniqueness is found, a new document authentication method based on the same principle can be developed.

To study the characteristics of print signatures, a representative test pattern was created as shown in Figure 6.2(a). The pattern comprises of four rounded dots. The diameter of the dots is $\frac{1}{360}$ inches and the horizontal and vertical distance between two adjacent dots is $\frac{1}{240}$ inches. These two numbers are selected by taking both the physical limitations of the printer and experimental results into consideration. The size of the dots is larger than the theoretically smallest dots the printer can print (in this case $\frac{1}{600}$ inches for a 600 dpi printer), so that the dots are clearly visible after printing. On the other hand, the dots are enough small for the random misplacement of toner powder to be significantly noticeable around their boundaries. The distance between two adjacent dots and the configuration of dots ensures that the printed dots will not merge together, which is very useful for a segmentation process. The number of dots balances the authentication performance and required computational resources [2].
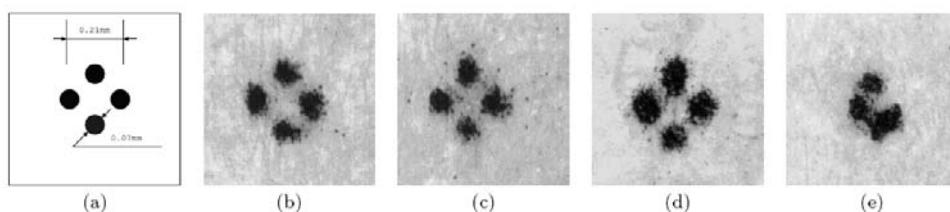
Figure 6.2: Printouts and photocopies of the testing pattern [2] [3].

Figure 6.2 shows some experimental printouts and photocopies examined under a 200× microscope. Image (b) and (c) are the test pattern printed using HP2 LaserJet 8100 (600dpi) office printer. Image (d) is the same test pattern printed on a high resolution HP LaserJet 4050 (1200 dpi) printer. Apparently, the dissimilarity among these patterns is large. Even for the same printer, a large variance is obtained. Image

(e) is a photocopy of image (b) using a 600×600 dpi digital photocopier Minolta Di152f3. It is quite obvious that the photocopied image is very different from the original one.

Besides the test pattern, occurrences of random toner powder misplacement can also be noticed at boundaries of printed characters, as shown in Figure 6.3, where images (a-e) are the source character, two test printouts on LaserJet 8100, one test printout on LaserJet 4050, and a photocopy of (b) on the Minolta photocopier respectively. The same phenomenon is noticed as with the previous experiment that the print signature is random and non-repeatable for each print run.
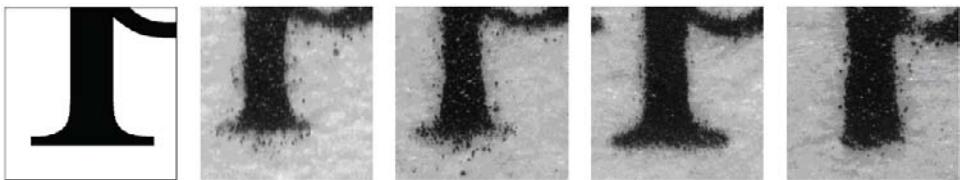
Figure 6.3: Printouts and photocopies of character "p" [2] [3].

Many such experiments have been performed and have consistently observed this occurrence for several types of laser printers. The experiments demonstrate the uniqueness and randomness of the proposed print signature. This method utilizes some features of this phenomenon to authenticate the originality of printer paper documents.

Without loss of generality, one proposed method is described based on the type of print signature shown in Figure 6.3. The test pattern used in the experiment is known as the secure pattern as it enables certain security features. With some minor modification, the method can also apply to the print signature detected on printed characters as well as hand signatures.

As illustrated in Figure 6.4, this method contains two procedures: registration and authentication.
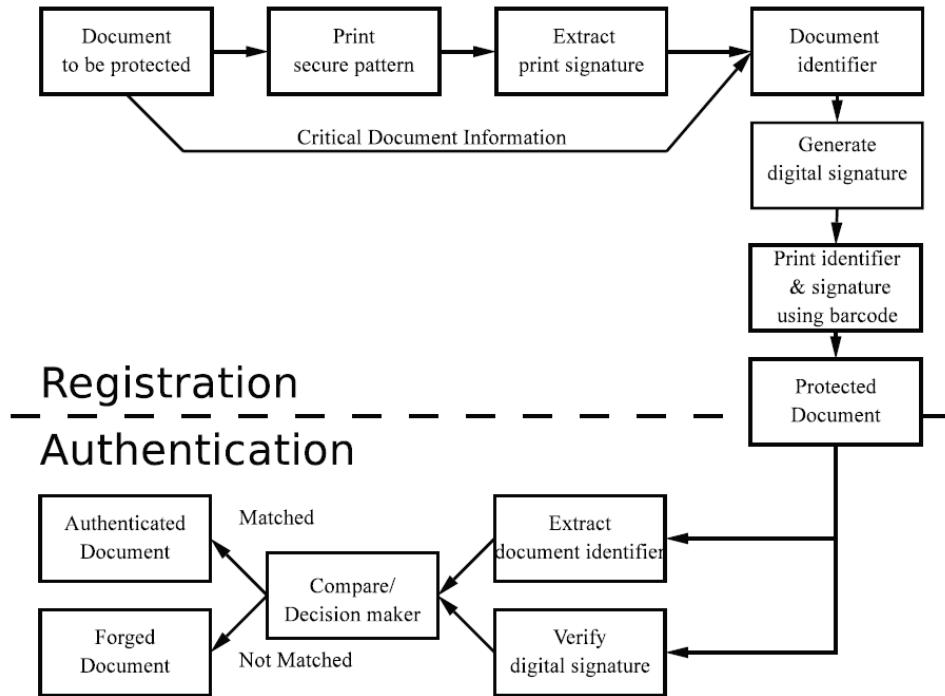
Figure 6.4: System design.

Registration: Given a document to be protected, the secure pattern is printed onto some blank area of the paper. Several auxiliary landmarks are also printed around the pattern to facilitate alignment. The printed paper is then examined by a microscope. Features describing the print signature such as the shape of the dots are detected and extracted. The feature description, together with some critical information about the document (such as the seat number in a concert ticket), forms an unique identifier for this specific document and specific print run. A digital signature is then generated for the identifier. The digital signature and the identifier are printed onto the same document using digital encoding methods such as bar codes or OCR fonts. These printed information and the secure pattern are used for later authentication.

Authentication: In order to verify the authenticity and originality of a printed document, feature extraction is performed, like in the registration process to get the feature description of the print signature. Also, the encoded information is read from the document using either a bar code or an OCR scanner. The digital signature is verified first to ensure there has been no modifications to the document identifier. The extracted feature is compared with the decrypted one, and the decrypted critical document information with the contents on the paper, through a decision process. If the results match, the document is considered to be authentic

and original. Otherwise it considered to be a fake or to have been tampered with.

As the laser printing technology improves, the printing resolution will become even higher. However, as long as the underlying mechanism is unchanged, it is still expected to see the random phenomenon on each copy of printed paper.

This will only entail the use of microscopes of even higher resolution. This method can be readily extended to other document types such as offset-printed documents, ink-jet printed documents, or manually signed documents. It basically reduces to the task of finding unique random phenomenon in each copy of the document to be used as a signature. For example, the ink trail for each manually signed document is unique. As long as the uniqueness is found, a new document authentication method based on the same principle can be developed.

6.3   Document Authentication with RSE

RSE (Render Sequence Encoding) [89], as the name indicates, encodes hidden data into the layout information of the formatted documents by modifying the display sequences of words or characters, without changing the content or appearance of the document. The encoding is achieved by specific permutation of display sequence, and the specific permutation is further linked to a Hamiltonian tour in the Exact Travel of Salesman Problem (XTSP). It needs a properly chosen cost matrix for the XTSP, and then the cost of the tour is made modulo equal to the content of digest. Because of the hardness of XTSP, with published, one cannot forge the signing process and attack the RSE authentication method. Unlike the existing digital watermarking methods which hide information into pure text documents or image based documents, RSE hides information into formatted documents. Those formatted documents contain both text data and layout information. Typical ones are PostScript (PS®1), Portable Document Format (PDF), Printer Control Language (PCL®2), and Device Independent Document (DVI). Formatted documents combines the advantages of both a text document and an image based document. They have small file sizes and platform-neutral page layout. It is widely used in electronic publishing, business, and administrative processing. It has merits of both digital signatures and digital watermarking: like digital signature, with verification key/information, one can perform verification, but not (or with great difficulties to) reverse engineer the signing process; it provides content protection, and survives transcoding processes.

Based on the RSE watermarking scheme, RSE authentication method adopts modular XTSP to authenticate the document. The security of RSE authentication method is guaranteed by the intractability of XTSP. The advantage of RSE authentication method over digital signature is its small authenticator size. With this

feature RSE authentication is adaptable to very short documents. Another feature of RSE authentication is its compatibility with most popular document formats. It can be integrated into existing systems with only minor changes at the creation and rendering ends of the whole workflow. RSE authentication method facilitates the "management of rights holders' relationship" by establishing trust among parties involved in document exchange. It can thus be a major building block in the whole DRM system for electronic documents.

6.4 Passive Image Authentication: Passive-Blind Image Forensics (PBIF)

However, for PBIF, no prior information is needed, as indicated by the image forensics process [90]. There are various works on image forgery detection that are based on the imaging process authenticity. For a digital camera, the scene radiance goes through the camera lens before being captured by an array of imaging sensors. The camera lens often has the optical low-pass effect for anti-aliasing. The imaging sensors are spatially allocated for measuring three types of colored light. To produce a three-color image, the missing color needs to be interpolated by de-mosaicing. In order to ensure that a white point in the image scene is rendered as white in the final image, the color is adjusted through white-balancing. The image may go through enhancements such as the contrast, sharpness, and saturation adjustment. Finally, gamma correction is applied for dynamic range compression and pleasing visual effects.

Natural image statistics (NIS) in the wavelet domain has been used for the forensic verification purpose. They show experiments for distinguishing authentic images from a few other types of images, i.e. stego images (images containing hidden messages), computer graphics, and print-and-scan images. In ongoing work, there are investigations of a number of other NIS, such as NIS in the power spectrum domain, NIS in the spatial local image patch, and NIS in higher-order statistics, primarily for distinguishing photographic images and computer graphics. To distinguish images captured by different cameras based on their physical device characteristics, one can use camera response function and fixed pattern noise. The former allows one to distinguish different models of camera and the latter different cameras.

Image post-processing clues raise suspicion for image forgery and help image forgery detection. Higher-order wavelets statistics are used for detecting image print-and-scan and steganography. Image operations, such as resampling, JPEG compression, and adding of noise, are modelled as linear operators and estimated by linear image deconvolution. Double JPEG compression has been given special attention in the PBIF literature. It is observed that double JPEG compression results in a periodic

pattern in the JPEG DCT coefficient histogram. Based on this observation, an automatic system that performs image forensics is developed. It is also found that the distribution of the first digit of the JPEG DCT coefficients can be used to distinguish a singly JPEG compressed image from a doubly compressed one. JPEG quantization tables for cameras and image editing software are shown to be different and may serve as a useful forensics clue. As in the case for an image, double MPEG compression artifacts are also observed when a video sequence is modified and MPEG re-encoded. Finally, there are also works on detecting duplicated image fragments due to the copy and paste operation.

As it is the goal for a forger to fool the forensic system, a forger can gather information on the forensic system and post-process the forgery so that it escapes the forensic system. Such an action is identified as a forensic system attack. Studying the potential forger's attack on a forensic method is necessary before one can design a counter-attack strategy. Apart from the counter-attack measure proposed for the recaptured attack described, the current work is very limited in this aspect. Indeed, there are many possible types of attacks on a forensic system. Below, three major types are discussed, i.e., oracle attack, recapturing attack, and post-processing attack.

Once a forgery creator has an unlimited access to a forgery detector, an oracle attack can be launched. The forger can incrementally modify the forgery guided by the detection results until it passes the detector with a minimal visual quality loss. In order to make the task of estimating the detection boundary more difficult, the work proposes a method of converting a parametric decision boundary into a fractal (non-parametric) one, so that an accurate estimation of the boundary requires a much larger number of sample points on the decision boundary. The oracle attack issue is addressed by modifying the temporal behavior of the detector such that the duration for returning a decision is lengthened when observing a sequence of similar-content input images, which is the hallmark of an oracle attack. The delay strategy can be designed so that the total time needed for an oracle attack to succeed is painfully long.

Apart from the protocol level attack, forgers could apply various post-processing operations to mask image forgery artifacts. This problem can be addressed by the post-processing detection techniques mentioned before. Furthermore, heavy post processing is often needed to mask the forgery artifacts.

A more sophisticated post-processing approach would be to simulate the device signature so that the forgery has a consistent device signature. However, such an attack is difficult to implement in practice as the simulated device signature has to be strong enough to mask the inconsistency in the first device signature, and thus results in a tremendous image quality loss.

An attacker can also produce a seemingly authentic image or video by recapturing the sound and sight produced from an image or a video. For example, an image can be printed out and recaptured by a camera. However, such an attack is difficult in practice, as to produce a good quality recaptured duplicate, a subtle and complicated setup for rendering the realistic sound and sight is needed, which not always feasible. For example, a printed image may contain the perceivable halftoning artifacts and its 2D fatness may lacks certain 3D visual effects. Furthermore, recapturing does not remove all the inconsistencies in an image or a video, which is particularly obvious in the scene inconsistencies.

6.5   Multimedia Forensics

Undoubtedly, multimedia forensics has become an extremely hot research area, especially with the exponential growth of digital camera technology and rapid use of the Internet. From the famous "flat tiger" to the "scandal" due to photo leaking, someone persuades the public to accept the photographs, someone hastens people to ignore the widely spread photos. However, genius is not fake, everything is traceable. In this section, a focus is on the emerging issue of and identifying of imagery from a computing point of view. The evidence is collected from the well known and accepted collections in real situations and use the evidence to support the original viewpoint. The multimedia materials are taken advantage of in a community or environment and encourage researchers from multiple disciplines to join in the study of this emerging application problem.

In photo forensics, one of important evidences is the EXIF of a digital photo. Figure 6.5 provides the EXIF entities that can be extracted from the JPEG file header so far.

Duplication detection is very helpful in image forensics. This is because the duplication points can not be removed from the images completely. The interesting points for duplicate detection are computed by using a block-based correlation approach,

Image Description	Scene Capture Type=	GPS Longitude=
Make	Sharpness=	GPS Altitude Ref=
Model	Unknown tag (0xC4A5)=	GPS Altitude=
Orientation	Compression=	GPS Time-Stamp=
X Resolution	Thumbnail Offset=	GPS Satellites=
Y Resolution	Thumbnail Length=	GPS Status=
Resolution Unit	Thumbnail Data=	GPS Measure Mode=
Date/Time	Index=	GPS DOP=
YCbCr Positioning	Version=	GPS Speed Ref=
Copyright	GPS Version ID=	GPS Speed=
Exposure Time	GPS Latitude Ref=	GPS Track Ref=
F-Number	Light Source=	GPS Track=
Exposure Program	Scene Capture Type=	GPS Img Direction Ref=
ISO Speed Ratings	Sharpness=	GPS Img Direction=
Exif Version	Unknown tag (0xC4A5)=	GPS Map Datum=
Date/Time Original	Compression=	GPS Dest Bearing Ref=
Date/Time Digitized	Thumbnail Offset=	GPS Dest Bearing=
Configuration	Thumbnail Length=	GPS Dest Distance Ref=
Compressed Bits Per Pixel	Thumbnail Data=	GPS Dest Distance=
Aperture Value	GPS Latitude=	
Light Source=	GPS Longitude Ref=	

Figure 6.5: EXIF entities of JPEG file header.

and incorporating a novel image area saliency measure in the computations. The duplicate decision is largely based on the types of images that are being compared. Therefore, in the third stage, the detection to change areas using class-specific similarity metrics adopts three distinct approaches: (1) Direct comparison of visual features extracted from the two images; (2) Classification of changes into a limited set of object areas; (3) Application of specialized object detectors. Finally, detectors for image areas are learnt from the training using the visual appearance and the final duplicate decision is made through statistical inference based on the outputs of the detectors, global image classifiers, and class-specific similarity metrics.

RANSAC (Random Sample Consensus) is an effective data-driven alignment and verification technique, it only generates a limited number of hypotheses with a sampling process guided by the matching of feature attributes, it has been used for alignment based verification for object recognition. The RANSAC techniques iterate through all the models in the database and therefore are linear in the number of models in the database.

The Harris corner detector is one of the main tools in interest points selection, correspondence weighting function is employed to search for the most likely match in the secondary image. For each candidate location, a local motion optimization using the KLT method is performed [91]. After trying to improve each correspondence, the algorithm re-computes the regression predictions and repeats the pair wise correspondence optimization. Termination occurs when an interaction completes without making further improvement. After finding a set of high likelihood correspondences, the local weighted regression method is used to interpolate the offset vectors, obtaining a dense corresponding field.

SIFT features have been used for duplication detection between images as these are

Fundamentals of Media SecurityDigital Multimedia Authentication and Forensics

invariant across a substantial range of affine distortion. From SIFT, points can be searched and matched in order to get a confidence level between two images. These detection points, m_1 and m_2, correspond to the salient points, the matching points are m, thereafter, the confidence between two photos are: $p(m_1, m_2) = \frac{m_1 m_2}{m_2}$.

6.5.1 Media Forensics Using Biometrics

Another technique used within image forensics, specifically for palmprint recognition, employs Fisher Linear Discriminative Analysis (FLDA) and Gabor filter banks [92]. This technique takes a palmprint and convolves it with a bank of Gabor filters at different scales and rotations in order to achieve robust palmprint feature extraction. After extraction, FLDA is applied for dimensionality reduction and class separability.

A general 2-D Gabor function $g(x, y)$ is defined as:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + j2\pi Wx\right] \tag{6.1}$$

where the spatial coordinates (x, y) denote the centroid localization of an elliptical Gaussian window and W is the frequency of a sinusoidal plane wave along the X-axis. The parameters σ_x and σ_y are the space constants of the Gaussian envelop along x and y axes, respectively. The Fourier transform $G(u, v)$ of the Gabor function $g(x, y)$ can be written as:

$$G(u, v) = \exp\left\{-\frac{1}{2}\left[\frac{(u-f)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\} \tag{6.2}$$

where f represents the frequency of the sinusoidal plane along the horizontal axis and the frequency components in the x and y direction are denoted by the pair (u, v), while $\sigma_u = \frac{1}{2\pi\sigma_x}$ and $\sigma_v = \frac{1}{2\pi\sigma_y}$. By considering a non-orthogonal basis set formed by Gabor functions, a localized frequency description can be obtained by expanding a signal with this basis. Self-similar class functions, known as Gabor Wavelets, can be generated by dilations and rotations of the mother wavelet $g(x, y)$ through the generating function:

$$g_{mn}(x, y) = a^{-m}g(x', y'), a > 1 \tag{6.3}$$

by considering $m = 1, \ldots, L$ and $n = 1, \ldots, K$. L and K denote the total number of dilations and orientations, respectively, and:

Download free eBooks at bookboon.com

$$
\begin{aligned}
x' &= a^{-m}(x\cos\theta + y\sin\theta) \\
y' &= a^{-m}(-x\sin\theta + y\cos\theta)
\end{aligned}
\tag{6.4}
$$

where $\theta = \frac{n\pi}{K}$ is the angle. To ensure that the energy is independent of m, a scale factor a^{-m} is introduced. By considering the redundant information presented in the filtered images due to the non-orthogonality of the Gabor wavelets, Manjunath [93] designed a strategy to reduce the redundancy of the GWFB. This strategy aims to maintain the half-peak magnitude of the filter responses touches each other in the frequency spectrum.

Let U_l and U_h denote the lower and the upper center frequencies of interest, respectively. The design strategy results in the following equations for computing the filter parameters σ_u and σ_v [93].

$$
a = \left(\frac{U_h}{U_l}\right)^{\frac{-1}{S-1}}
\tag{6.5}
$$

$$
\sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\ln 2}}
\tag{6.6}
$$

$$
\sigma_v = \tan\left(\frac{\pi}{2k}\right)\left[U_h - 2\ln\left(\frac{\sigma_u^2}{U_h}\right)\right]\left[2\ln 2 - \frac{(2\ln 2)^2\sigma_u^2}{U_h^2}\right]^{\frac{-1}{2}}
\tag{6.7}
$$

where $f = U_h$. In order to eliminate the sensitivity of filter responses to absolute intensity values the real components of 2D Gabor filters are biased by adding a constant to make them with zero mean.

As previously mentioned, Gabor representation of a palmprint image $i(x, y)$ can be obtained by convolving the image with the family of Gabor filter as follows:

$$I_{m,n}(x, y) = \iint i(x, y) g_{mn}^*(x - x_0, y - y_0) dx_0 dy_0 \qquad (6.8)$$

where $I_{m,n}(x, y)$ denotes the result corresponding to the Gabor filter at scale L and orientation K and $*$ indicate the complex conjugate.

The next problem that must be overcome involves localizing and segmenting a palmprints ROI (Region of Interest). A ROI should include all the important patterns from a palmprint's texture, such as principle lines and wrinkles. Typically, this method operates on a fixed square region inside the palm. However, this region is not proportional to the palm size and is thus unsuitable to extract all necessary palmprint information from complete lines and wrinkles.

To address this problem, a number of image processing techniques are employed in order to improve ROI extraction. The techniques include image binarization, which involved a Gaussian low pass filter, which smoothes the image, the image then has a thresholding method applied to it[94]. Alignment is another key point which must be considered. Two robust key points are computed on the image and these are used to determine the amount of rotation required. Finally, the image is at a stage whereby it has been suitable processed so that the ROI can be determined or extracted.

After the ROI has been determined and extracted, processing can begin upon that specific image. These regions of interest can be cross referenced against a database of previous palmprints using techniques previously mentioned, such as SIFT, in order to determine who the palmprints may belong to. This type of analysis can prove to be very useful for many forensic applications, particularly at crime scenes.

6.6 Summary

This chapter, mainly discussed the hot issues about media authentication and forensics that are currently being researched in this field. Video authentication and general document authentication was also discussed. Also discussed was the issues of media forensics along with some solutions. This emerging area of media authentication and forensics is currently moving at a rapid pace with many researchers devoting a lot of time and effort into proving whether images and videos are authentic. Along with this, forensics of a biometric nature were also discussed. Using different techniques to segment and analyze palmprints can be very important when it comes to determining who left the prints. The majority of these techniques use many standard image processing methods which already have a wide variety of literature on them.

Bibliography

[1] R. Nishimura, "Information hiding into interaural phase differences for stereo audio signals," *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 1189 –1192, Sep. 2009.

[2] B. Zhu, J. Wu, and M. S. Kankanhalli, "Print signatures for document authentication," in *ACM Conference on Computer and Communications Security*, 2003, pp. 145–154.

[3] S. Huang and J. K. Wu, "Optical watermark," Patent 7 366 301, April, 2008. [Online]. Available: http://www.freepatentsonline.com/7366301.html

[4] M. Ju, S. Kim, and T.-H. Kim, "A study on digital media security by hopfield neural network," in *ISNN '07: Proceedings of the 4th international symposium on Neural Networks*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 140–146.

[5] E. Praun, H. Hoppe, and A. Finkelstein, "Robust mesh watermarking," in *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 49–56.

[6] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 417–424.

[7] K. Castleman, *Digital Image Processing*. Prentice Hall, 1979.

[8] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 4, pp. 379–423, 623–656, July, October 1948.

[9] J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," in *Proceedings of the First International Workshop on Information Hiding*. London, UK: Springer-Verlag, 1996, pp. 207–226.

[10] J. Zöllner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke, and G. Wolf, "Modeling the security of steganographic systems," in *Proceedings of the Second International Workshop on Information Hiding.* London, UK: Springer-Verlag, 1998, pp. 344–354. [Online]. Available: http://portal.acm.org/citation.cfm?id=647595.731539

[11] C. Cachin, "An information-theoretic model for steganography," *Inf. Comput.*, vol. 192, no. 1, pp. 41–56, 2004.

[12] S. Katzenbeisser and F. A. Petitcolas, Eds., *Information Hiding Techniques for Steganography and Digital Watermarking.* Norwood, MA, USA: Artech House, Inc., 2000.

[13] G. Zhu and S. Zhang, "Image information hiding design and implementation," *International Symposium on Information Engineering and Electronic Commerce*, pp. 130 –134, May. 2009.

[14] J. Fridrich, "Minimizing the embedding impact in steganography," in *MM&Sec '06: Proceedings of the 8th workshop on Multimedia and security.* New York, NY, USA: ACM, Sep. 2006, pp. 2–10.

[15] C. Manikopoulos, Y.-Q. Shi, S. Song, Z. Zhang, Z. Ni, and D. Zou, "Detection of block DCT-based steganography in gray-scale images," *IEEE Workshop on Multimedia Signal Processing*, pp. 355 – 358, Dec. 2002.

[16] W. Ella, "Detecting steganography on a large scale," *Crossroads*, vol. 15, no. 2, pp. 3–6, 2008.

[17] J. Ren, Y. Xia, and Z. Ma, "Information hiding algorithm based on predictive coding," *International Conference on Information Technology and Computer Science*, vol. 2, pp. 11 –14, Jul. 2009.

[18] D. L. Parnas, *The secret history of information hiding.* New York, NY, USA: Springer-Verlag New York, Inc., 2002.

[19] Z. Xu, H.-F. Ling, F. Zou, Z. Lu, and P. Li, "Robust image copy detection using multi-resolution histogram," in *Proceedings of the 11th ACM SIGMM International Conference on Multimedia Information Retrieval, MIR 2010*, Mar. 2010, pp. 129–136.

[20] F. Zou, Z. Lu, H.-F. Ling, and Y. Yu, "Real-time video watermarking based on extended m-sequences," in *Proceedings of the 2006 IEEE International Conference on Multimedia and Expo, ICME 2006*, Jul. 2006, pp. 1561–1564.

[21] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673 –1687, Dec. 1997.

[22] W. Liu, L. Dong, and W. Zeng, "Optimum detection for spread-spectrum watermarking that employs self-masking," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 645 –654, Dec. 2007.

[23] O.-C. Chen and W.-C. Wu, "Highly robust, secure, and perceptual-quality echo hiding scheme," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 629 –638, Mar. 2008.

[24] R. J. Anderson, Ed., *First International Workshop on Information Hiding, Cambridge, U.K., May 30 - June 1, 1996, Proceedings*, ser. Lecture Notes in Computer Science, vol. 1174. Springer, 1996.

[25] R. K. Ward, Ed., *Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, Sep. 2000*, vol. 3, 2000.

[26] M. Dorairangaswamy, "Protecting digital-image copyrights: A robust and blind watermarking scheme," in *First International Conference on Networked Digital Technologies*, Jul. 2009, pp. 423 –428.

[27] D. Zhang, B. Wu, J. Sun, and H. Huang, "A new robust watermarking algorithm based on DWT," in *2nd International Congress on Image and Signal Processing*, Oct. 2009, pp. 1 –6.

[28] R. Ridzon and D. Levicky, "Robust digital watermarking in DFT and LPM domain," in *ELMAR, 2008. 50th International Symposium*, vol. 2, Sep. 2008, pp. 651 –654.

[29] H. Li, Z. Qin, and L. Shao, "Audio watermarking pre-process algorithm," in *IEEE International Conference on e-Business Engineering*, Oct. 2009, pp. 165–170.

[30] L. Wang, S. Emmanuel, and M. S. Kankanhalli, "EMD and psychoacoustic model based watermarking for audio," in *IEEE International Conference on Multimedia and Expo*, Jul. 2010, pp. 1427 –1432.

[31] R. Petrovic and D. Yang, "Audio watermarking in compressed domain," in *9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services*, Oct. 2009, pp. 395–401.

[32] K. Raghavendra and K. Chetan, "A blind and robust watermarking scheme with scrambled watermark for video authentication," in *IEEE International Conference on Internet Multimedia Services Architecture and Applications*, Dec. 2009, pp. 1–6.

[33] A. Essaouabi and E. Ibnelhaj, "A 3D wavelet-based method for digital video watermarking," in *First International Conference on Networked Digital Technologies*, Jul. 2009, pp. 429 –434.

[34] P. Horvatic, J. Zhao, and N. J. Thorwirth, "Robust audio watermarking based on secure spread spectrum and auditory perception model," in *Proceedings of the IFIP TC11 Fifteenth Annual Working Conference on Information Security for Global Information Infrastructures*. Deventer, The Netherlands, The Netherlands: Kluwer, B.V., 2000, pp. 181–190.

[35] W. Zhu, Z. Xiong, and Y.-Q. Zhang, "Multiresolution watermarking for images and video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 545 –550, Jun. 1999.

[36] O. Benedens, "Geometry-based watermarking of 3d models," *Computer Graphics and Applications, IEEE*, vol. 19, no. 1, pp. 46–55, Jan. 1999.

[37] R. Ohbuchi, H. Masuda, and M. Aono, "Watermarking three-dimensional polygonal models through geometric and topological modifications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 551–560, May. 1998.

[38] M. Yeung and B.-L. Yeo, "Fragile watermarking of three-dimensional objects," in *Proceedings of the International Conference on Image Processing*, vol. 2, Oct. 1998, pp. 442 –446.

[39] Y. Song and T. Tan, "Comparison of four different digital watermarking techniques," in *5th International Conference on Signal Processing Proceedings*, vol. 2, 2000, pp. 946–950.

[40] R. van Schyndel, A. Tirkel, and C. Osborne, "A digital watermark," in *IEEE ICIP '94*, vol. 2, Nov. 1994, pp. 86 –90.

[41] R. Wolfgang and E. Delp, "A watermark for digital images," in *Proceedings of the International Conference on Image Processing*, vol. 3, Sep. 1996, pp. 219–222.

[42] I. Pitas, "A method for signature casting on digital images," in *International Conference on Image Processing*, vol. 3, Sep 1996, pp. 215–218.

[43] Y.-C. Hou and P.-M. Chen, "An asymmetric watermarking scheme based on visual cryptography," *WCCC-ICSP 5th International Conference on Signal Processing Proceedings*, vol. 2, pp. 992–995, 2000.

[44] D. Tzovaras, N. Karagiannis, and M. G. Strintzis, "Robust image watermarking in the subband and discrete cosine transform domain," in *EUSIPCO*, Rhodes, Sep. 1998.

[45] R. Wolfgang, C. Podilchuk, and E. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1108 –1126, Jul. 1999.

[46] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3-4, pp. 313–336, 1996.

[47] F. Hartung and B. Girod, "Digital watermarking of MPEG-2 coded video in the bitstream domain," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, Apr. 1997, pp. 2621 –2624.

[48] M. Swanson, B. Zhu, B. Chau, and A. Tewfik, "Object-based transparent video watermarking," in *IEEE First Workshop on Multimedia Signal Processing*, Jun. 1997, pp. 369 –374.

[49] J. Zhang, A. T. S. Ho, G. Qiu, and P. Marziliano, "Robust video watermarking of H.264/AVC," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 2, pp. 205 –209, Feb. 2007.

[50] L. Joyeux, O. Buisson, B. Besserer, and S. Boukir, "Detection and removal of line scratches in motion picture films," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, p. 553, 1999.

[51] D. P. Huttenlocher, G. A. Klanderman, and W. A. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[52] D. P. Huttenlocher, R. H. Lilien, and C. F. Olson, "Approximate Hausdorff matching using eigenspaces," in *Proceedings of the ARPA Image Understanding Workshop*, 1996, pp. 1181–1186.

[53] E. Belogay, C. Cabrelli, U. Molter, and R. Shonkwiler, "Calculating the Hausdorff distance between curves," in *Information Processing Letters*, vol. 64, 1997, pp. 17–22.

[54] C. Olson, "A probabilistic formulation for Hausdorff matching," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 1998, pp. 150 –156.

[55] V. Mangulis, "Security of a popular scrambling scheme for tv pictures," in *RCA Review*, 1980, pp. 423–432.

[56] Y. Matias and A. Shamir, "A video scrambling technique based on space filling curves," in *CRYPTO '87: A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology*.   London, UK: Springer-Verlag, 1988, pp. 398–417.

[57] W. Ding, W. Yan, and D.-X. Qi, "A novel digital image hiding technology based on tangram and conways' game," in *International Conference on Image Processing*, vol. 1, 2000, pp. 601–604.

[58] S. Hojjat, K. Sadri, and S. Shirani, "Multiple description coding of audio using phase scrambling," in *IEEE International Conference on Multimedia and Expo*, Apr. 2008, pp. 153 –156.

[59] H. Li and Z. Qin, "Audio scrambling algorithm based on variable dimension space," in *International Conference on Industrial and Information Systems*, Apr. 2009, pp. 316–319.

[60] J. Herre and E. Allamanche, "Compatible scrambling of compressed audio," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999, pp. 27–30.

[61] S. Borujeni, "Speech encryption based on fast fourier transform permutation," in *The 7th IEEE International Conference on Electronics, Circuits and Systems*, vol. 1, 2000, pp. 290–293.

[62] D. Qi, J. Zou, and X. Han, "A new class of scrambling transformation and its application in the image information covering," *Science in China Series E: Technological Sciences*, vol. 43, pp. 304–312, 2000, 10.1007/BF02916835. [Online]. Available: http://dx.doi.org/10.1007/BF02916835

[63] D. Qiu, J. Lu, X. Sun, and J. Jiang, "Application of graphic minimal covering algorithm in the distribution of surveillance cameras in small and medium-sized city road networks," in *International Conference on Computer Design and Applications (ICCDA)*, vol. 1, Jun. 2010, pp. 200–203.

[64] W. K. Wong, J. Liew, C. K. Loo, and W. K. Wong, "Omnidirectional surveillance system for digital home security," in *International Conference on Signal Acquisition and Processing*, Apr. 2009, pp. 8–12.

[65] C.-H. Liu and C.-C. Fan, "The design of remote surveillance system for digital family," in *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Sep. 2009, pp. 238–241.

[66] L. Nguyen and L. Vu, "Secured wireless digital video surveillance for distributed enterprises," in *IEEE 60th Vehicular Technology Conference*, vol. 7, sep. 2004, pp. 5330 – 5334 Vol. 7.

[67] W. Yan and M. S. Kankanhalli, "Detection and removal of lighting & shaking artifacts in home videos," in *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*. New York, NY, USA: ACM, 2002, pp. 107–116.

[68] B. C. Kuo and F. Golnaraghi, *Automatic Control Systems*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[69] R. Jain, "Experiential computing," *Communications of the ACM*, vol. 46, no. 7, pp. 48–55, 2003.

[70] ——, "Out-of-the box data engineering - events in heterogeneous environments," in *ICDE*, U. Dayal, K. Ramamritham, and T. M. Vijayaraman, Eds. IEEE Computer Society, 2003, pp. 8–21.

[71] J. Wang and M. S. Kankanhalli, "Experience based sampling technique for multimedia analysis," in *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*. New York, NY, USA: ACM, 2003, pp. 319–322.

[72] M. Kankanhalli, J. Wang, and R. Jain, "Experiential sampling in multimedia systems," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 937–946, Oct. 2006.

[73] A. Rosenfeld, "Picture processing by computer," *ACM Computing Surveys*, vol. 1, no. 3, pp. 147–176, 1969.

[74] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, "Event detection and analysis from video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 873–889, Aug. 2001.

[75] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*. CRC Press, 2001. [Online]. Available: http://www.cacr.math.uwaterloo.ca/hac/

[76] N. Degara-Quintela and F. Perez-Gonzalez, "Visible encryption: using paper as a secure channel," in *Security and Watermarking of Multimedia Contents V*, E. J. D. III and P. W. Wong, Eds., vol. 5020, no. 1. SPIE, 2003, pp. 413–422. [Online]. Available: http://link.aip.org/link/?PSI/5020/413/1

[77] A. J. Sellen and R. H. Harper, *The Myth of the Paperless Office*. Cambridge, MA, USA: MIT Press, 2003.

[78] R. L. van Renesse, *Optical Document Security, 3rd edition*. Boston/London: Artech House, 2005.

[79] E. Zeira and D. Ellett, "Verification methods employing thermally–imageable substrates," Patent 6 107 244, August, 2000. [Online]. Available: http://www.freepatentsonline.com/6107244.html

[80] N. R. Wagner, "Fingerprinting," *IEEE Symposium on Security and Privacy*, vol. 0, p. 18, 1983.

[81] A. D. Narasimhalu, W. Wang, and M. S. Kankanhalli, "Method for utilizing medium nonuniformities to minimize unauthorized duplication of digital information," Patent 5 412 718, May, 1995. [Online]. Available: http://www.freepatentsonline.com/5412718.html

[82] J. Brassil, "Tracing the source of a shredded document," in *IH '02: Revised Papers from the 5th International Workshop on Information Hiding*. London, UK: Springer-Verlag, 2003, pp. 387–399.

[83] E. Metois, P. Yarin, N. Salzman, and J. R. Smith, "Fibre-fingerprint identification," in *Third Workshop on Automatic Identification*, March 2002.

[84] T. Pavlidis, J. Swartz, and Y. P. Wang, "Fundamentals of bar code information theory," *Computer*, vol. 23, no. 4, pp. 74–86, 1990.

[85] ——, "Information encoding with two-dimensional bar codes," *Computer*, vol. 25, no. 6, pp. 18–28, 1992.

[86] G. B. Rhoads, "Identification/authentication system using robust, distributed coding," Patent 5 745 604, April, 1998. [Online]. Available: http://www.freepatentsonline.com/5745604.html

[87] M. Naor and A. Shamir, "Visual cryptography," in *Proc. of EuroCrypt'94*, vol. LNCS:950, no. 1-12. Berlin: Springer-Verlag, 1994.

[88] A. Shamir, "Method and apparatus for protecting visual information with printed cryptographic watermarks," Patent 5 488 664, January, 1996. [Online]. Available: http://www.freepatentsonline.com/5488664.html

[89] V. Mandolkar, "RSE for electronic text document protection," in *2nd International Conference on Computer Engineering and Technology*, vol. 1, April 2010, pp. 39–43.

[90] W. Wang, J. Dong, and T. Tan, "A survey of passive image tampering detection," in *IWDW '09: Proceedings of the 8th International Workshop on Digital Watermarking*.   Berlin, Heidelberg: Springer-Verlag, 2009, pp. 308–322.

[91] T. Nguyen and D. Taubman, "Optimal linear detector for spread spectrum based multidimensional signal watermarking," in *16th IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 113–116.

[92] M. Laadjel, A. Bouridane, F. Kurugollu, O. Nibouche, and W. Yan, "Partial palmprint matching using invariant local minutiae descriptors," *Transactions on Data Hiding and Multimedia Security*, vol. 5, pp. 1–17, 2010.

[93] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 837–842, 1996.

[94] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, January 1979.